DISCLAIMER:

This document does not meet the
current format guidelines of
the Graduate School at
The University of Texas at Austin.

It has been published for
informational use only.

The Report committee for Mark Arjun Bond

Certifies that this is the approved version of the following Report:

Using the Filter-forward backward sampling algorithm in second-order Bayesian latent growth modeling

APPROVED BY

SUPERVISING COMMITTEE:

_____

Susan N. Beretvas, Supervisor

_____

Tiffany Whittaker

Using the Filter-forward backward sampling algorithm in second-order Bayesian latent growth modeling

by

Mark Arjun Bond, B.S.Psy., B.A.

Report

Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of
Master or Science in Statistics

The University of Texas at Austin
December 2016

Using the Filter-forward backward sampling algorithm in second-order Bayesian latent growth modeling

by

Mark Arjun Bond, M. S. Stat.

The University of Texas at Austin, 2016

SUPERVISOR: Susan N. Beretvas

In educational and social science research, large-scale testing data are frequently collected longitudinally so that researchers can evaluate change over time. Researchers may then wish to assess the impact of various explanatory variables on student growth in achievement outcomes. Use of structural equation modeling allows for the modeling of item-level measurement error and allows growth trajectories to vary by student. If a Bayesian perspective is adopted, one may use psychometric information known a priori about the test items in the estimation process, which may improve ability estimation. In addition, Bayesian estimation procedures, like the Kalman filter, are able to take advantage of the autoregressive structure of time series data to obtain closed-form solutions for ability distributions. In contrast, a structural equation modeling-based approach using likelihood-based estimation would need to rely on iteratively updating proposed model estimates and checking a discrepancy function, which might achieve a local minimum, or fail to converge. Researchers have previously estimated second-order latent growth models with IRT elements, and this work will expand upon that literature in a number of ways. Bayesian research to date has typically relied on use of the WinBUGS software program to estimate these models which does not allow for certain distributional assumptions. For instance, although certain non-informative priors may be specified, it is not possible to use improper non-informative priors

with WinBUGS. Also, WinBUGS does not take advantage of the autoregressive structure of a time series analysis to speed up the estimation process, which is possible using the Kalman filter. Because thousands of iterations of calculation and random-number generation are recommended when using a Bayesian Gibbs sampler, the improved computational efficiency of the Kalman filter may make growth models easier to estimate. When time series data are highly correlated, the Kalman filter, theoretically, should improve the rate of convergence for a Gibbs sampler. Furthermore, research on second-order latent growth modeling has not evaluated the use of informative priors for item parameters. The present work will address these limitations. Parameters based on educational psychology research will be used to simulate a dataset which will be analyzed with and without the Kalman filter. Then, convergence diagnostics, including the traceplot, will be assessed to determine whether the Kalman filter improved the rate of convergence. Additionally, both informative and non-informative priors will be used for item parameters, and parameter recovery will be assessed.

Table of Contents

Introduction

Social scientists are often interested in understanding change over time. Within the context of educational psychology, the process of learning is one form of change over time. When student learning is carefully measured and properly modeled, it is possible to address a wide variety of research questions. On the most basic level, is the learning process exponential, quadratic, or linear in nature? More practically speaking, why do some students learn more than others? What can be changed to help students achieve?

Researchers can investigate these questions with latent growth modeling, a particular form of structural equation modeling (Bollen & Curran, 2006). In the simplest latent growth model, observed scores are modeled as a function of a random intercept term and a random slope term. In other words, the model can be used to provide an initial ability and growth term estimate for each person. Many of the above questions can be answered within this basic framework by adding explanatory variables to the growth term.

Recent research has focused more on what are called second-order latent growth models. In these models, multiple items are measured at each time point, and the measurement error in the items are separated from the true latent scores which are modeled as the direct indicators of the growth model's parameters. Simulation studies have indicated that second-order latent growth modeling has a number of advantages, including a greater probability to detect statistically

significant differences between growth terms (von Oerzen, Hertzog, Lindenberger, & Ghisletta, 2010), better estimates of item reliability (Geiser, Keller and Lockhart, 2013), and the ability to test whether the measurement model changes over time (Ferrer, Balluerka, and Widaman, 2008).

It is possible that, even with a measurement component, the basic framework of latent growth modeling as described may not adequately model change over time. Specifically, this form of latent growth modeling does not take into account how observations that are adjacent in time may be related to one another. It is possible that achievement at a particular time point might be correlated with adjacent achievement scores above and beyond the association modeled between initial ability and student growth. Simulation studies have suggested that failing to correctly model autocorrelation over and above the associations captured by the linear growth parameter can negatively influence results by resulting in greater type 1 error rates when testing whether the growth term is nonzero, along with reducing power for detecting nonzero growth (Murphy & Pituch, 2009; Kwok, West, & Green, 2007).

Most recently, Murphy, Beretvas, and Pituch (2013) conducted a simulation study evaluating the estimation of an autoregressive second-order latent growth model. Findings indicated that although modeling an autoregressive trend reduced relative bias, variance terms remained substantially biased. Furthermore when an autoregressive model was estimated, between 14.3 and 43.7% of the models failed to converge.

To attempt to address these limitations, the present work seeks to expand upon Murphy et al., (2013) in a number of ways. First, a Bayesian estimation method is proposed. In this estimation method, a particular Gibbs sampler, the Kalman filter, will account for the autocorrelated nature of the indicators of the latent constructs. Unlike the methods employed in Murphy et al. (2013), this Gibbs sampler is theoretically guaranteed to converge eventually (Roberts & Smith, 1994). A simulation study is proposed wherein the use of the forward-filter backward sampling algorithm is evaluated.

Literature review

This literature review will cover four main topics. First, the mathematical and statistical specification of first-order latent growth modeling is introduced. The next section presents the second-order latent growth model along with results from simulation studies that have assessed estimation of the model. A variety of time series approaches will be reviewed. Finally, Bayesian methods for time series analysis, particularly the Kalman filter, will be introduced.

First Order latent growth modeling

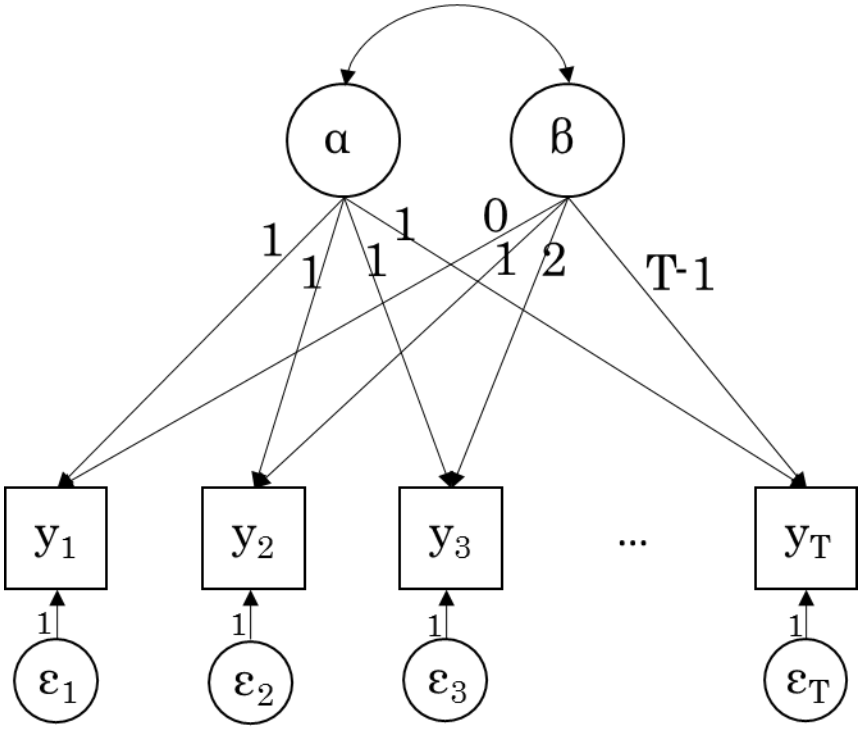
In educational psychology, it sometimes happens that large-scale testing data are collected longitudinally so that researchers can evaluate change over time. Researchers may then wish to assess the impact of various explanatory variables on student growth in achievement outcomes. Although growth over time may be analyzed using analysis of variance (ANOVA) or conventional multiple regression, doing so assumes that all students grow at the same rate. Another form of analysis, latent growth models using structural equation modeling (SEM) or multilevel modeling allows growth trajectories to vary by student (see, for example, Hancock, Harring, & Lawrence, 2013). Using the SEM framework, the model is mathematically specified in the following way (Bollen & Curran, 2006):

$$\boldsymbol{y} = \boldsymbol{\Lambda\eta} + \boldsymbol{\varepsilon} \qquad (1)$$

where $\boldsymbol{y}$ is the $T \times 1$ vector of observed responses for student i, $\boldsymbol{\eta}$ is the $m \times 1$ vector of m factors for student i, $\boldsymbol{\Lambda}$ is the $T \times m$ matrix of factor loadings, and $\boldsymbol{\varepsilon}$ is the $T \times 1$ vector of residuals, which are typically assumed i.i.d. Normal with a mean of zero.

Supposing that growth in a latent construct is hypothesized to be parameterized using only an intercept term and a linear growth term, m equals two, and the first column of $\mathbf{\Lambda}$ contains the value $\mathbf{1}$, whereas the second column contains the time variable coded as follows $\mathbf{t} = (0, 1, 2, \ldots, T-1)^t$, where ^{t} denotes the transpose of that row vector. Also in this example, we denote the first element of $\boldsymbol{\eta}$ as α and the second element as β, which are, respectively, the intercept and growth terms. A structural-regression diagram (explained in, e.g., Kline, 2011) for this running example is provided in Figure 1, below.

Figure 1. A structural-regression diagram of a linear unconditional unidimensional first-order latent growth model



The observation, y_t, for each time point t *is related to the covarying random intercept and slope factors α and ß. Observations at each time point have a separate random error term,* ε_t.

Next, consider how exactly α and ß in $\boldsymbol{\eta}$ are specified as being random. They may be expressed as function of an average and a deviation from the average:

$$\boldsymbol{\eta} = \boldsymbol{\mu}_{\eta} + \boldsymbol{\zeta} \qquad (2)$$

where the $m \times 1$ mean vector $\boldsymbol{\mu}_{\eta}$ contains elements μ_{α} and $\mu_{ß}$, and the $m \times 1$ residual vector $\boldsymbol{\zeta}$ containing ζ_{α} and $\zeta_{ß,}$ which are assumed to be normally distributed with

means of zero. It is assumed that $\boldsymbol{\zeta}$ and $\boldsymbol{\varepsilon}$ are independent of each other, however, the elements of $\boldsymbol{\zeta}$ are not typically assumed to be independent of each other, so their $m \times m$ covariance matrix $\boldsymbol{\Psi}$ is given by:

$$\boldsymbol{\Psi} = \begin{bmatrix} \psi_{\alpha\alpha} & \psi_{\alpha\beta} \\ \psi_{\alpha\beta} & \psi_{\beta\beta} \end{bmatrix} \qquad (3)$$

where variance terms, $\psi_{\alpha\alpha}$ and $\psi_{\beta\beta}$, correspond to the intercept and slope terms, respectively, and $\psi_{\alpha\beta}$ represents the covariance between the intercept and slope factors.

The covariance matrix of the latent factors may be used to compute the variance of the observed responses. Plugging Equation 2 into Equation 1 gives:

$$\boldsymbol{y} = \boldsymbol{\Lambda}(\boldsymbol{\mu}_{\boldsymbol{\eta}} + \boldsymbol{\zeta}) + \boldsymbol{\varepsilon} \qquad (4)$$

meaning that the model-implied covariance matrix $\sum$ can be expressed as

$$\sum = \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}^{t} + \boldsymbol{\Theta}\varepsilon \qquad (5)$$

where $\boldsymbol{\Theta}\varepsilon$ is a $T \times T$ diagonal matrix of variance terms for $\boldsymbol{\varepsilon}$ at each time point.

An HLM specification of the latent growth model is theoretically identical to the SEM specification (Curran, 2003). In fact, structural equation modeling subsumes certain kinds of hierarchical linear modeling in the sense that many different kinds of multi-level structures may be represented in an SEM (e.g., Muthén, & Asparouhov, 2011, Curran, 2003). However, a unidimensional linear latent growth model may be estimated in SEM using only a covariance matrix and a

mean vector, whereas estimating these same models using conventional HLM software requires the raw data (Curran, 2003). Beyond concerns about convenience, it is more conventional in the literature to use SEM to model measurement error (contrast, e.g., Schumacker & Lomax, 2010, with Gelman & Hill, 2007). Since this project ultimately aims to incorporate a measurement-error model, the SEM notation will be used.

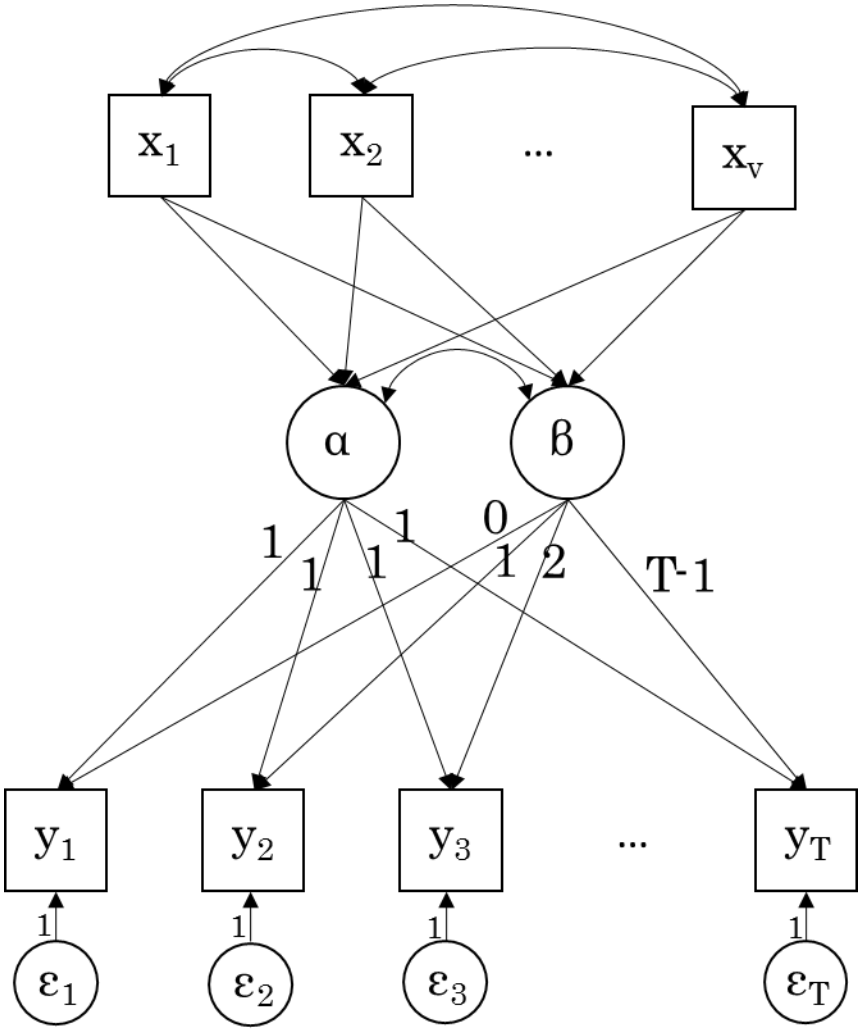
Regardless of the parameterization employed, an unconditional LGM may be of somewhat limited use to social science researchers. That is, an unconditional LGM merely describes growth in the relevant variables and variances in the growth trajectory parameters. Under the definition of "unconditional", there are no predictor variables modeled as explaining variability in growth parameters. Researchers may be more interested in the reasons why, for example, certain students learn more quickly than others, rather than a description of students' average growth. To explain the variance in the latent intercept and growth terms, the conditional latent growth model is introduced.

Consider again a unidimensional linear latent growth model, this time with a $v \times 1$ vector of explanatory variables, $\boldsymbol{x}$, with, for instance, variables that are called $x_1, x_2, \ldots, x_v$. For the sake of argument, suppose that all v variables were collected prior to obtaining the first outcome measure. A structural-path model in Figure 2, below. Then, keeping the notation the same as in the above example, once again, the observed outcomes are specified as in Equation 1, above. However, at level two, the latent vector $\boldsymbol{\eta}$ is modeled conditionally:

$$\boldsymbol{\eta} = \boldsymbol{\Gamma x} + \boldsymbol{\zeta} \qquad (6)$$

where $\boldsymbol{\Gamma}$ is an $m \times v$ matrix where each row represents regression weights for the m^{th} latent factor. If, as is typically the case, the first vector in $\boldsymbol{x}$ contains a 1 for each student, then the first column of $\boldsymbol{\Gamma}$ contains the conditional intercept term, which, in this example, would represent the α and β terms when all the other elements of $\boldsymbol{x}$ are set to zero. Once again it is assumed that $\boldsymbol{\zeta}$ and $\boldsymbol{\varepsilon}$ are independent of each other; additionally, it is now assumed that $\boldsymbol{\zeta}$ and $\boldsymbol{x}$ are independent of each other. Again, the covariance matrix $\boldsymbol{\Psi}$ for the elements of $\boldsymbol{\zeta}$ need not be diagonal, and is specified in Equation 3. Of course, in a conditional model, the elements of $\boldsymbol{\Psi}$ represent the variances and covariances that are not explained by the terms in $\boldsymbol{x}$.

Figure 2. A structural-regression diagram of a linear unidimensional first-order latent growth model, where the intercept and slope are conditional on predictors x_1 through x_v.



Recall that, for the sake of this example, the explanatory variables x_1, x_2, …, x_v were assumed to have been collected prior to testing. Also, in this example, the growth term ß remains the same at each time point for a given participant. It may be that researchers are interested in a more dynamic system in which, for instance,

growth varies from time point to time point, as a result of changing independent variables. A more dynamic model would be more appropriate when a great deal of time-sensitive data are collected, and when it is straightforward to determine which variables are independent of the measurement process. However, for the sake of simplicity, the present work focuses on growth models in which the growth term is fixed for a particular person.

Once a conditional model is fit, researchers may evaluate the matrix $\boldsymbol{\Gamma}$ to assess the degree to which particular predictor variable are related to initial student ability and student growth. Whereas unconditional latent growth modeling merely provides a description of student growth, it is possible for conditional latent growth modeling to regress those growth terms on student characteristics.

Both of the above models, which were diagrammed in Figures 1 and 2, may be limited since a linear growth term is assumed. However, nonlinear terms may be easily incorporated. Suppose that researchers wish to estimate an intercept, a linear growth term, and a quadratic growth term. Then, $\boldsymbol{\eta}$ may be expanded to include α, ß, and an extra term, say, δ, meaning that m, the number of dimensions describing the trajectory, is equal to three instead of two. One may then specify the observed outcomes as in Equation 1, above, with the third column of $\boldsymbol{\Lambda}$ written as the vector $(0^2, 1^2, 2^2, \ldots, T-1^2)^t$. In other words, to estimate a squared growth term, researchers can just square each element of $\mathbf{t}$, (which was defined under Equation 1), and add another element of $\boldsymbol{\eta}$. Any polynomial growth term may be specified in this way, provided that there are enough time points for that polynomial to be

identified (Bollen & Curran, 2006). In a similar fashion, logarithmic or exponential growth terms may be specified by applying analogously appropriate functions to each element of $\mathbf{t}$ (e.g., Grimm, Ram, & Hamagami, 2011; Grimm & Ram, 2009). Beyond the augmentation of $\boldsymbol{\eta}$ with additional terms and that of $\boldsymbol{\Lambda}$ with additional columns, both conditional and unconditional nonlinear latent growth models may be specified using the same equations as above.

Even with explanatory variables and nonlinear growth terms, latent growth modeling as described above is limited in a number of ways. First, note that the observed responses for each student at each time point are modeled as a single summary measure. In other words, in this simplest growth model, a total or average score is modeled for the outcome rather than item score by item score. Thus, the measurement error of each specific item is subsumed into a single error term for an observed outcome. Also, another kind of error, which could be present, is unmodeled in a first-order latent growth model. Possibly, a student's latent ability at time *t* could depend on circumstances that are particular to the time point *t.* For example, a measurement of ability at a specific time could be error-prone due to a flawed test administration procedure during that time. In other words, there may be time-specific error sources for each ability estimate that could impact the estimation of student ability which are unrelated to measurement error from the items. However, in first-order latent growth modeling, time-specific error and measurement error are combined into one term. Properly separating these terms would allow a more accurate apportioning of measurement error.

Another potential limitation arises from the way that these models are typically estimated. First-order latent growth models may be estimated using maximum likelihood-based estimation, relying on iteratively updating proposed model estimates and checking a discrepancy function (Bollen & Curran, 2006, pp. 39-42). Unfortunately, this local discrepancy function may achieve a local minimum; also, the process of iteratively updating a likelihood-based estimate may fail to converge. Latent growth models may also be estimated using Bayesian methods, such as a Gibbs sampler (e.g., Elliot, Gallo, Ten Have, Bogner, & Katz, 2005), which are mathematically guaranteed to converge eventually (Roberts & Smith, 1994). Gibbs samplers are discussed more thoroughly in the section describing the Kalman filter. However, researchers have used a likelihood-based methodology in a wide variety of applications, which are discussed in the next section.

A great deal of research has been done using structural equation models that are similar to the one diagrammed in Figure 2, above. That is, although researchers often incorporate nonlinear techniques in fitting a latent growth model, it is often the case that researchers do not estimate a latent ability term for every student at each time point. For example, one study used a unidimensional latent growth model to estimate the effect of parental support and monitoring on adolescent alcohol use (Barnes, Reifman, Farrell, & Dintcheff, 2000). The two latent variables, as shown in Figure 2, were baseline alcohol misuse (analogous to α) and the slope in this term (analogous to ß). Notably, the vector $\boldsymbol{y}$ in this article was a composite measure, taking the average of three separate item scores. Then, the measurement error $\boldsymbol{\varepsilon}$ in

some way represents the measurement error of this composite term. Possibly it would be of use to estimate the measurement error of specific items also, since researchers could then assess whether or not those items are worth including. This composite error term also confounds measurement error with time effects; a particular time point may have more or less error for reasons beyond measurement error. For instance, in the first wave of data collection, the participants were between 13 to 16 years of age, and in the last wave of data collection, participants were between 18 to 22 years old. Because some of the students may have not been living with their parents in the last wave, one might argue that some participants had a greater incentive to lie about their alcohol use in the first wave, relative to the last wave. This is error from an effect that may change over time, that is, error due to a time effect. This is theoretically distinct from those measurement errors that are not associated with a time component. However, in Barnes et al. (2000), those two effects are confounded in the $\boldsymbol{\varepsilon}$ term. Moreover, it is assumed that the measurement error has the same variance across all time points, which may or may not be realistic, depending on the topic of study. Finally, when items are parceled, it is assumed that the averaged items are unidimensional and that they are all equally indicative of the latent construct (Little, Rhemtulla, Gibson, & Schoemann, 2013). If the items are not actually unidimensional, there is a threat to construct validity which cannot be evaluated.

To summarize, when researchers fail to specify an exact measurement model for the observed outcomes, (the $\boldsymbol{y}$s), measurement error and time-specific error are

confounded. Measurement error is assumed to be homoscedastic, which may or may not be appropriate, depending on the area of study. Also, it is assumed that the items are unidimensional. Unfortunately, a good amount of applied research has been done in this way, where the observed outcome is some composite measure, and the psychometric properties of the individual items are not statistically modeled (see, e.g., Fleming, Mason, Mazza, Abbott, & Catalano, 2008; Colder et al., 2001; Barnes et. al, 2000; Chan & Schmitt, 2000; Duncan, Duncan, Biglan, & Ary, 1998). All this research may have been better served to model the measurement error of each item separately, with each item loading onto a latent measure of ability at each time point. That is the approach of what is called second-order latent growth modeling.

Second-Order Latent Growth Modeling

In second-order latent growth modeling, there is a latent measure of a trait for each participant at each time point, modeling the measurement error of each item individually (See, e.g., Harring, Kohli, Silverman, & Speece, 2012; Sayer & Cumsille, 2001). Suppose that, for each time point, every participant responds to K items. Then let $\boldsymbol{y}$ be the $(K*T)\times 1$ vector with elements defined so that the first element is the response to first item at the first time point, the second element is the response to the second item at the first time point, and so on. Call the elements of this vector y_{11}, y_{12}, …, y_{1K}, y_{2K}, …, and y_{TK}. Also, let $\boldsymbol{\eta}$ be a $(T*m)\times 1$ vector, whose first element is the value of the first latent variable at the first time point, whose second element is the value of the first latent variable at the second time point, and

so on; call the elements of this vector $\boldsymbol{\eta}_{11}$, $\boldsymbol{\eta}_{12}$, …, $\boldsymbol{\eta}_{1T}$, $\boldsymbol{\eta}_{21}$, …, and ζ_{Tm_2} Furthermore, let $\boldsymbol{\eta_1}$ be the vector formed by the first T elements of $\boldsymbol{\eta}$, let $\boldsymbol{\eta_2}$ be the vector formed by the next T elements of $\boldsymbol{\eta}$, and so on. Note that the quantity T*m represents the total number of latent constructs at level one. Then the measurement model may be written identically to Equation 1, but with $\boldsymbol{\Lambda}$ written as a $(K*T)\times(T*m)$ matrix and $\boldsymbol{\varepsilon}$ written as a $(K*T)\times 1$ vector instead. Similarly to $\boldsymbol{y}$, above, call the elements of $\boldsymbol{\varepsilon}$ ε_{11}, ε_{12}, …, ε_{1K}, ε_{2K}, …, and ε_{TK}. Each row of $\boldsymbol{\Lambda}$ gives the factor loadings for a particular item at a particular time point. The matrix $\boldsymbol{\Lambda}$ is assumed to be block-diagonal; in other words, trait estimates at a particular time point do not load onto items at another time point. If it is assumed that the measurement model is the same across time points, then all the blocks that make up $\boldsymbol{\Lambda}$ are equal to the $K\times m$ submatrix $\boldsymbol{\Lambda_\alpha}$, formed by taking only the first K rows and first m columns of $\boldsymbol{\Lambda}$. Supposing that the measurement equation is written in this way, the structural portion may be written similarly to Equation 8 (Hancock, Kuo, & Lawrence, 2001):

$$\boldsymbol{\eta} = \boldsymbol{\Gamma\xi} + \boldsymbol{\zeta} \qquad (7)$$

with $\boldsymbol{\zeta}$ an $(T*m_2)\times 1$ error vector that is similar to the same term in Equation 8, with $\boldsymbol{\Gamma}$ an $(T*m)\times m_2$ matrix of factor loadings, and with $\boldsymbol{\xi}$ an $m_2\times 1$ vector of second-order latent factors; note that the number of second order latent factors, m_2, need not be the same as the number of first-order latent factors, m. Typically, models estimate fewer second-order factors than first-order factors (see, e.g., Hancock, et al., 2001). Similarly to $\boldsymbol{\eta}$, call the elements of $\boldsymbol{\zeta}$ ζ_{11}, ζ_{12}, …, ζ_{1T}, ζ_{21}, …,

and $\square_{m_2 T}$; furthermore, let $\boldsymbol{\zeta_1}$ be the vector formed by the first T elements of $\boldsymbol{\zeta}$, let $\boldsymbol{\zeta_2}$ be the vector formed by the next T elements of $\boldsymbol{\zeta}$, and so on. Finally, the vector $\boldsymbol{\xi}$ is specified similarly to Equation 2:

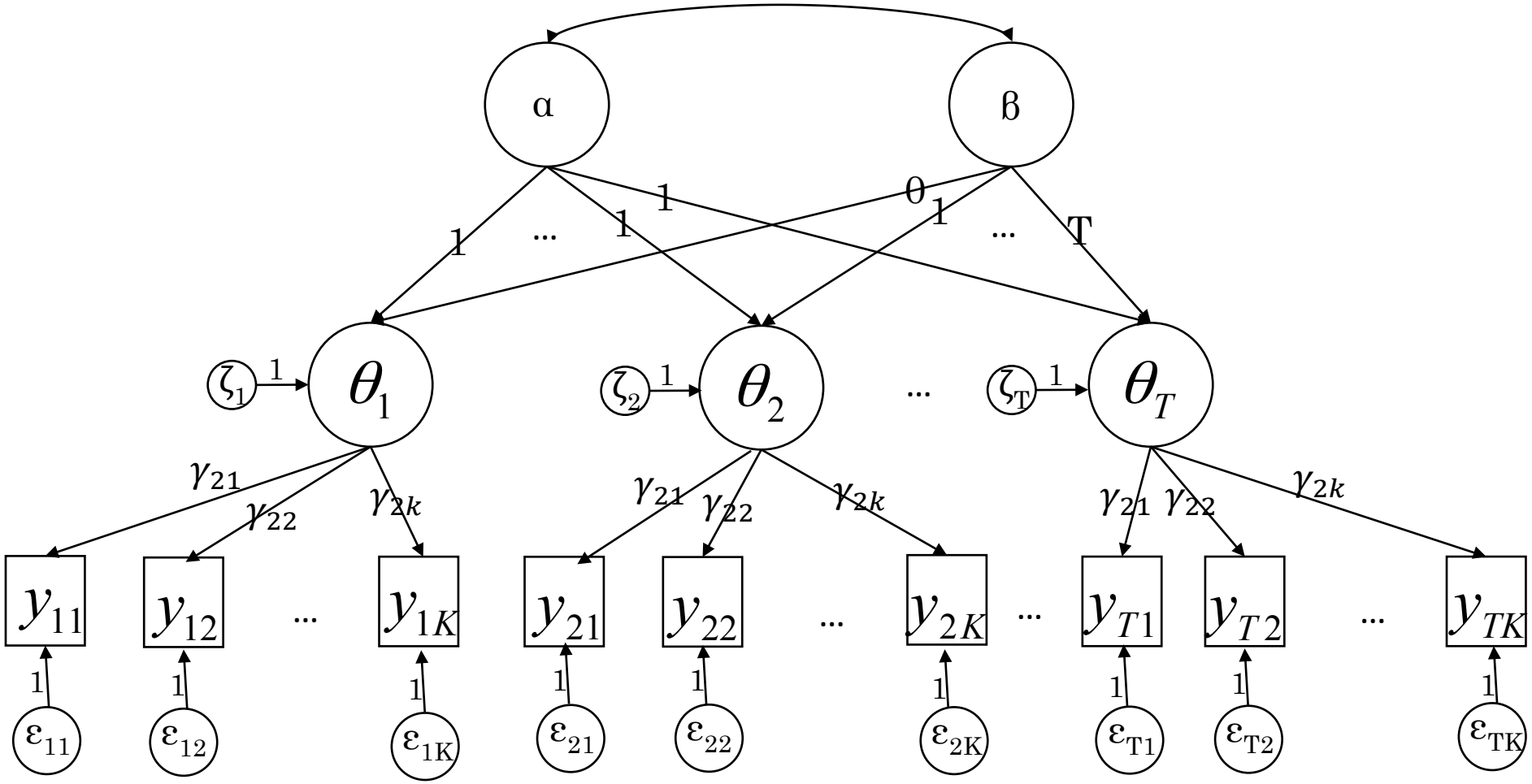
$$\boldsymbol{\xi} = \boldsymbol{\mu_\eta} + \boldsymbol{\zeta_\eta} \qquad (8)$$

with $\boldsymbol{\mu_\eta}$ an $m_2 \times 1$ mean vector and with $\boldsymbol{\zeta_\eta}$ the $m_2 \times 1$ disturbance vector representing the variance of the level-two latent factors, also assumed to normally distributed with means of zero.

For the sake of example, consider a linear unconditional unidimensional second-order latent growth model, assuming strong metric invariance across time points. A structural-regression diagram for such a model provided in Figure 3, below. Begin by considering Equation 10 in the context of this example. Since the model is assumed to be linear, unconditional, and unidimensional, let $m_2 = 3$, and allow the mean vector to consist of the elements 1, μ_α, and μ_β; then, let the elements of $\boldsymbol{\zeta_\eta}$ be 0, $\zeta_{\eta\alpha}$ and $\zeta_{\eta\beta}$. The quantities 1 and 0 in these vectors are set as fixed. Then, the vector $\boldsymbol{\xi}$ consists of the elements 1, α, and β; the latter two are, respectively, intercept and growth terms. Then, in Equation 9, the matrix $\boldsymbol{\Gamma}$ is block-diagonal, with blocks $\mathbf{1}$, $\mathbf{1}$, and $\mathbf{t}$, as defined in Equation 1. Computing the matrix algebra, $\boldsymbol{\eta_1} = \mathbf{1}$, and $\boldsymbol{\eta_2} = \alpha^*\mathbf{1} + \beta^*\mathbf{t}$; then, $\boldsymbol{\zeta_1} = 0$. For convenience, call the elements of $\boldsymbol{\eta_2}$ $\theta_1, \theta_2, \ldots, \theta_T$, and call the elements of $\boldsymbol{\zeta_2}$ $\zeta_1, \zeta_2, \ldots, \zeta_T$. Next, consider the measurement model as defined in Equation 1. Since strong metric invariance is assumed, it suffices to consider the $K \times m$ submatrix $\boldsymbol{\Lambda_\alpha}$. Call the respective columns of this matrix $\boldsymbol{\gamma_1}$ and $\boldsymbol{\gamma_2}$; call the

elements of these column vectors γ_{11}, γ_{12}, …, γ_{1K}, and γ_{21}, γ_{22}, …, and γ_{2K}, respectively. Computing the block-matrix algebra, then, the product vector on the right hand side of Equation 1 is equal to $\boldsymbol{\gamma_1 + \gamma_2 * \eta_2}$.

Figure 3. A structural-regression diagram of an unconditional linear unidimensional second-order latent growth model, assuming strong metric invariance across time points.



Recall that, typically, the error terms $\boldsymbol{\varepsilon}$ are assumed to be independent identically distributed Normal variables. If the vector $\mathbf{y}$ represents binary data, for example, if each element of $\mathbf{y}$ records whether or not an examinee correctly answered a question, then the assumption of normality would not be appropriate. Whereas normally distributed variables have support along the entire real line, in this case each element of $\boldsymbol{\varepsilon}$ would be restricted to lie between negative one and positive one. To avoid this inappropriate specification, a different link function for the vector $\mathbf{y}$ may be employed:

$$P(\mathbf{y}=1) = \Phi(\boldsymbol{\gamma}_1 + \boldsymbol{\gamma}_2 * \boldsymbol{\eta}_2) \qquad (9)$$

with $\Phi(x)$ representing the cumulative density function of the standard normal distribution. This specification may be thought of as an implementation of the Normal Ogive model in psychometrics (Lord, 1952; see also Fox & Glas, 2001).

Regardless of whether or not the vector $\mathbf{y}$ represents continuous or binary data, the second-order latent growth model has more variance terms than a first-order latent growth model. The error terms $\boldsymbol{\zeta}_2$ and $\square$ theoretically represent distinct forms of variation (Sayer & Cumsille, 2001). That is, the terms in $\boldsymbol{\zeta}_\eta$ represent variation in the initial state and the variation in the growth term; the terms in $\boldsymbol{\zeta}_2$ represent unaccounted variation in the latent factors $\boldsymbol{\theta}$. Also, for continuous $\mathbf{y}$ vectors, the terms in $\boldsymbol{\varepsilon}$ represent unaccounted variation in the vector of observed variables $\mathbf{y}$. In other words, the vector $\boldsymbol{\zeta}_2$ represents how the latent factor itself varies over time, independently of measurement error. In contrast, consider vector $\boldsymbol{\varepsilon}$ represented in figure 1.

For the sake of example, consider a latent growth model estimating the growth of mathematics achievement in grade-school children, assuming the data are Normally distributed. Suppose that, for the first testing occasion, the children happened to have unusually variable achievement, that is, achievement that is more variable than is observed at later time points. Suppose also that the questions asked in the first assessment were unusually poor at discriminating between high and low achieving students. In a first-order latent growth model, these two sources of variation, that is, the unaccounted variation in mathematics achievement over time and the variation due to measurement error, would both have to be contained in a single error term, ε_1, since that is the only possible random error associated with the first measurement, y_1. In this way, these two sources of variation are confounded in a single error term (Geiser, Keller, & Lockhart, 2013).

In contrast, a second-order latent growth model would not have this difficulty. Heteroskedastic mathematics scores would be captured in the variance of the term ζ_1, while poor measurement during the first testing would be modeled in the variance of the error terms (ε_{11}, ε_{12}, …, and ε_{1K}).

The direct measurement of unaccounted (or time-specific) factor variance allows for further statistical estimation and testing. For instance, consider the fitting of unconditional and conditional second-order latent growth models. Researchers may be interested in the proportion of variation in latent factors that is accounted for by conditioning those factors on exogenous variables. In second-order latent growth models, but not first-order latent growth models, this may be

accomplished by comparing the unexplained variance of the latent factors before and after these exogenous variables are included in the model. Moreover, in second-order latent growth models, researchers may test the null hypothesis that the unaccounted variation in the factors is equal to zero. A statistically significant result of this test would indicate that more variables may be necessary to fully account for the time-specific variation in the constructs. Again, this statistical test would not be possible for a first-order latent growth model, since factor variation is not explicitly included in that model.

Moreover, if this time-specific variance exists, then the reliability of the items is better estimated in a second-order latent growth model. Reliability in this context is defined as the variance of the latent variable divided by the total observed variance (Steyer, Ferring, & Schmitt, 1992). In Geiser, Keller and Lockharts' 2013 simulation study, it was shown mathematically that first-order latent growth models would, on average, tend to underrepresent the reliability of the items, given the existence of time-specific variance. Next, in the simulation study, differing amounts of occasion-specific variance were used to generate latent growth curves, then reliability was estimated using both first-order and second-order latent growth models. Finally, percentage bias terms were obtained by taking the difference of estimated and true reliability terms, dividing by the true reliability. When there was no time-specific variance, the first-order and second-order latent growth models were roughly comparable in terms of percentage bias. However, if there was even a small amount of occasion-specific variance, the first-order latent growth model had

between 12.1 to 29.9 percent bias, whereas the second-order latent growth model had only between 0.01 to 0.07 percent bias. This trend increased as occasion-specific variance increased, with a maximum percentage bias of 64.9 percent for first-order latent growth models, and a maximum percentage bias of 12.0 percent for second-order latent growth models. Thus, when there is a great deal of time-specific variance, there is a strong incentive to use second-order latent growth modeling rather than first-order latent growth modeling.

Suppose alternatively that there is relatively little time-specific variance when modeling latent growth. In this situation, it has been shown analytically (von Oerzen, Hertzog, Lindenberger, & Ghisletta, 2010) that second-order latent growth models have greater power to detect individual differences in growth than first-order latent growth models. In a latent growth modeling context, the power to detect individual differences in growth is the probability of correctly rejecting the null hypothesis that the variance in growth is equal to zero. Therefore, if the variability of this slope term is of substantive interest, a second-order latent growth model may be preferred. Moreover, a simulation study (von Oerzen et al., 2010) demonstrated that a larger number of observed variables increases the power to detect variability in intercept and slope terms, particularly when there is little time-specific variance. Since first-order latent growth models are unable to accommodate multiple observed variables at a single time point, this simulation study also provides evidence that second-order latent growth modeling is preferred when testing the variability of slope and intercept terms.

Second-order latent growth models may also allow for the factor loadings in Λ to differ, if needed. In other words, the measurement of the factor may be allowed to vary over time. It is also possible to statistically test for the need to do this (Ferrer, Balluerka, and Widaman, 2008). Since the measurement of the factor is not explicitly modeled in first-order latent growth modeling, such a procedure would not be possible using a first-order latent growth modeling. Indeed, first-order latent growth modeling must assume that the measurement of the factor remains stable over time (Ferrer et al., 2008). In an empirical research article, Ferrer, et. al (2008) showed that when these factor loadings actually differ, the results from first-order and second-order latent growth modeling provide widely disparate results for the growth term. In short, second-order latent growth modeling can test the assumption of equal factor measurement across time; when this assumption is violated and untested, the growth term may be biased. Again, unfortunately, first-order latent growth models are unable to be used to test this assumption, since they assume that factor measurement remains the same over time (Ferrer et al., 2008).

Second-order latent growth models are relatively recent to the educational psychology literature (Leite, 2007). Their specification may be intimidating to applied researchers, which may be one reason why methodologists note a paucity of second-order latent growth models in the educational psychology literature (Geiser, Keller, & Lockhart, 2013). However, similar models have a long history in the econometrics literature.

Time Series Approaches

Consider, for instance, the unidimensional autoregressive model with noise (Granger & Morris, 1976; further discussed in Box, Jenkins, & Reinsel, 2008; Chatfield, 2004; Soares, Gonclaves, & Gamerman, 2009; Petris, Petrone, & Campagnoli, 2009), which can be adapted to subsume a second-order latent growth model. Recall that the first level of a conditional first-order latent growth model is specified as

$$\boldsymbol{y} = \boldsymbol{\Lambda\eta} + \boldsymbol{\varepsilon} \qquad (10)$$

with $\boldsymbol{y}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\eta}$ and $\boldsymbol{\varepsilon}$ as specified in Equation 1, and that a conditional first-order latent growth model further models $\boldsymbol{\eta}$ as

$$\boldsymbol{\eta} = \boldsymbol{\Gamma x} + \boldsymbol{\zeta} \qquad (11)$$

with $\boldsymbol{\eta}$, $\boldsymbol{\Gamma}$, $\boldsymbol{x}$ and $\boldsymbol{\zeta}$ as specified in Equation 6. For a unidimensional autoregressive model with noise, index the vectors $\boldsymbol{y}$, $\boldsymbol{\eta}$, and $\boldsymbol{\varepsilon}$ by the time point t, so that the first level of the model is given by

$$\boldsymbol{y_t} = \boldsymbol{\Lambda\eta_t} + \boldsymbol{\varepsilon_t} \qquad (12)$$

for time-indexed data vector $\boldsymbol{y_t}$, time-indexed latent vector $\boldsymbol{\eta_t}$ and time-indexed error term $\boldsymbol{\varepsilon_t}$. Since $\boldsymbol{y_t}$ is the vector of observed responses at time t, it is a $K \times 1$ vector, where K is the number of items. Then, $\boldsymbol{\Lambda}$ is a $K \times m$ factor loading matrix, where m is the number of elements in the column vector $\boldsymbol{\eta_t}$, and, finally, let $\boldsymbol{\varepsilon_t}$ be a $K \times 1$ column vector. Then, the first column of the matrix contains item parameters b_1, b_2, …, b_k, and the second column contains discrimination parameters a_1, a_2, …, a_k. Then, further model $\boldsymbol{\eta_t}$ as

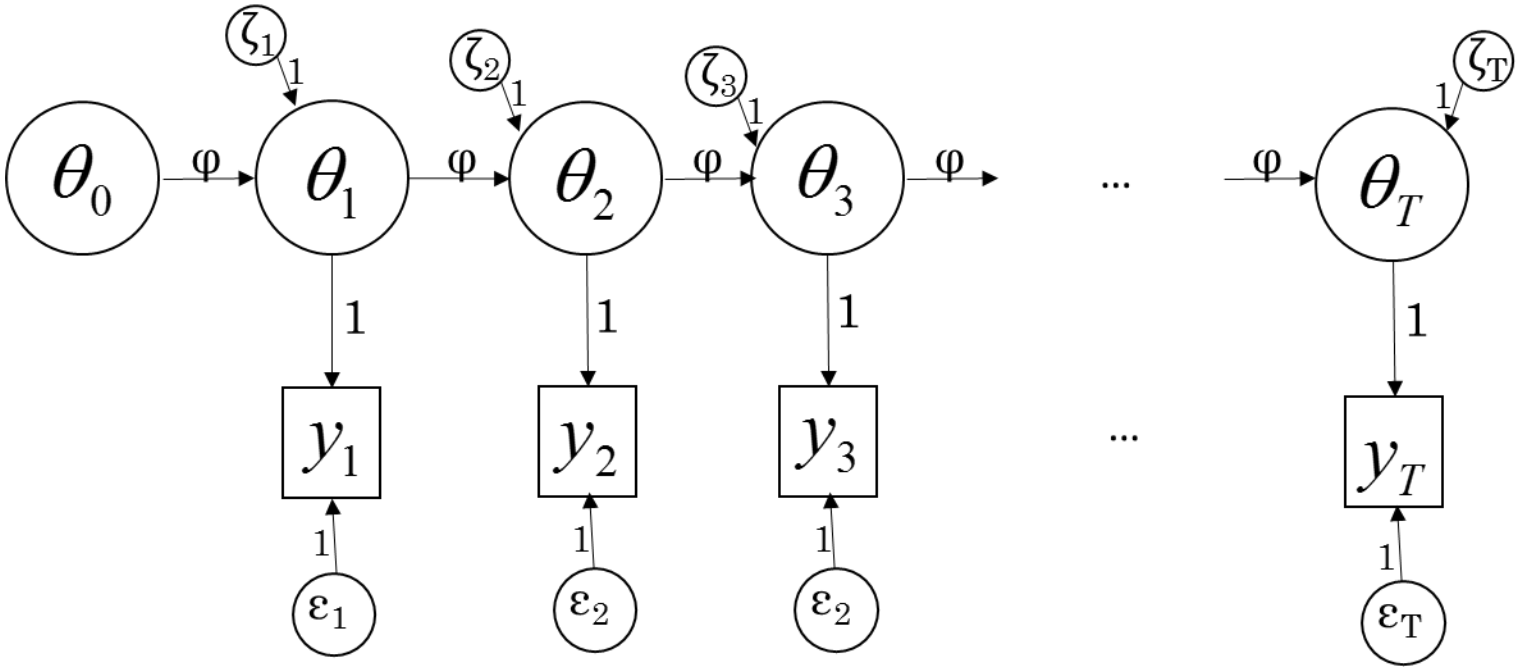
$$\boldsymbol{\eta_t} = \boldsymbol{\Gamma}\boldsymbol{\eta_{t-1}} + \boldsymbol{\zeta_t} \qquad (13)$$

with $\boldsymbol{\eta_t}$ and $\boldsymbol{\eta_{t-1}}$ as in Equation 12, above, with $\boldsymbol{\Gamma}$ as an $m \times m$ matrix of regression weights, and with $\boldsymbol{\zeta_t}$ as an $m \times 1$ error vector. This is called an "autoregressive" or "AR" model since each construct is regressed on an earlier value of that same construct. Furthermore, this particular Equation is called an AR(1) model, since $\boldsymbol{\eta_t}$ is only regressed on the one immediately previous value of $\boldsymbol{\eta}$; it would also be possible, for instance, to regress $\boldsymbol{\eta_t}$ on the two most recent values, which would be called an AR(2) model.

For the sake of this example, let m equal one, and denote the element of $\boldsymbol{\eta_t}$ as θ_t. Then, the matrix $\boldsymbol{\Gamma}$ has a single element; call that element φ. The parameter φ is known as the "autoregressive parameter", since it is a regression weight describing the impact of θ_t on itself at a later time. A structural-regression diagram for this example is provided in Figure 4, below.

To ensure that the variance of the θ estimates remains constant over time, a latent variable θ_0 is incorporated. Without this latent variable, the variance of θ_1 would equal some value, say, ζ, and the variance of all future time points would equal $\varphi^{2*}\zeta$. Incorporating the latent variable θ_0, with variance ζ, ensures that the variance of all the other latent variables equals $\varphi^{2*}\zeta$.

Figure 4. A structural-regression diagram of the unidimensional autoregressive time-series model with noise.



Like the first-order latent growth model, this autoregressive model may be modified in a variety of ways. For instance, rather than having a single observation at a time point, it is possible to include K observations y_{1t}, y_{2t}, …, y_{Kt} at each time point t, each with corresponding factor loadings γ_{21}, γ_{22}, …, γ_{2K}.

Employing this model necessitates the use of a number of assumptions. Similarly to the first-order latent growth model, in an autoregressive model it is typically assumed that the first order residuals $\boldsymbol{\varepsilon_t}$ are i.i.d., with means of zero, and are uncorrelated with each other at different time points (Hamilton, 1994). Also, it is typically assumed that the disturbances, $\boldsymbol{\zeta_t}$, are i.i.d., with means of zero, and are uncorrelated at different time points (Hamilton, 1994).

Additionally, it is commonly assumed that autoregressive time series models are at least weakly stationary. A time series process is said to be weakly stationary when two criteria are satisfied. The first criterion is that that the average value of the observed data remains the same across time points. The second criterion is that the covariance between two observed values that are *j* time points apart is the same across time points, for any value *j*. In other words, a time series process is weakly stationary when the means, variances, and covariances of the observed data remains the same across time shifts, given the model (Hamilton, 1994). For this assumption to be met, any time-dependent trend in the data must be correctly modeled.

This property is desirable for a variety of reasons. Firstly, long-term averages across all time points are mathematically guaranteed to converge to the population

mean for autoregressive time series models (Hamilton, 1994). Also, given weak stationarity, it is possible to forecast an autoregressive time series model arbitrarily far into the future (Hamilton, 1994), allowing researchers to predict how the observed data may change in the future. Finally, assuming weak stationarity for autoregressive models allows for autocovariances to be obtained analytically (Hamilton, 1994). In a time series process, the j^{th} autocovariance is defined as the covariance between an observed datum point y_t and an observed datum point $y_{t-j,}$ for a particular value of *j*. The autocovariance function takes as input that number *j* and outputs the corresponding autocovariance for a particular time series. The autocorrelation and autocorrelation function are similarly defined, dividing the corresponding autocovariance by the stationary variance of the observed data.

To satisfy the assumption of weak stationarity, it is necessary, but not sufficient, for the autoregressive parameter ϕ in Figure 4 to be less than one in absolute value. To see this, consider the following autoregressive process without noise. Let an observed datum point be a function of a previously observed datum point:

$$y_t = c + \varphi y_{t-1} + \varepsilon_t \qquad (14)$$

for y_t, y_{t-1}, φ, and ε_t as described above, and for c an intercept term. Suppose that this time series process is weakly stationary and recursively plug Equation 14 into itself. Then y_t may be expressed as

$$y_t = c + \varepsilon_t + \varphi(c + \varepsilon_{t-1}) + \varphi^2(c + \varepsilon_{t-2}) + \varphi^3(c + \varepsilon_{t-3}) + \cdots$$

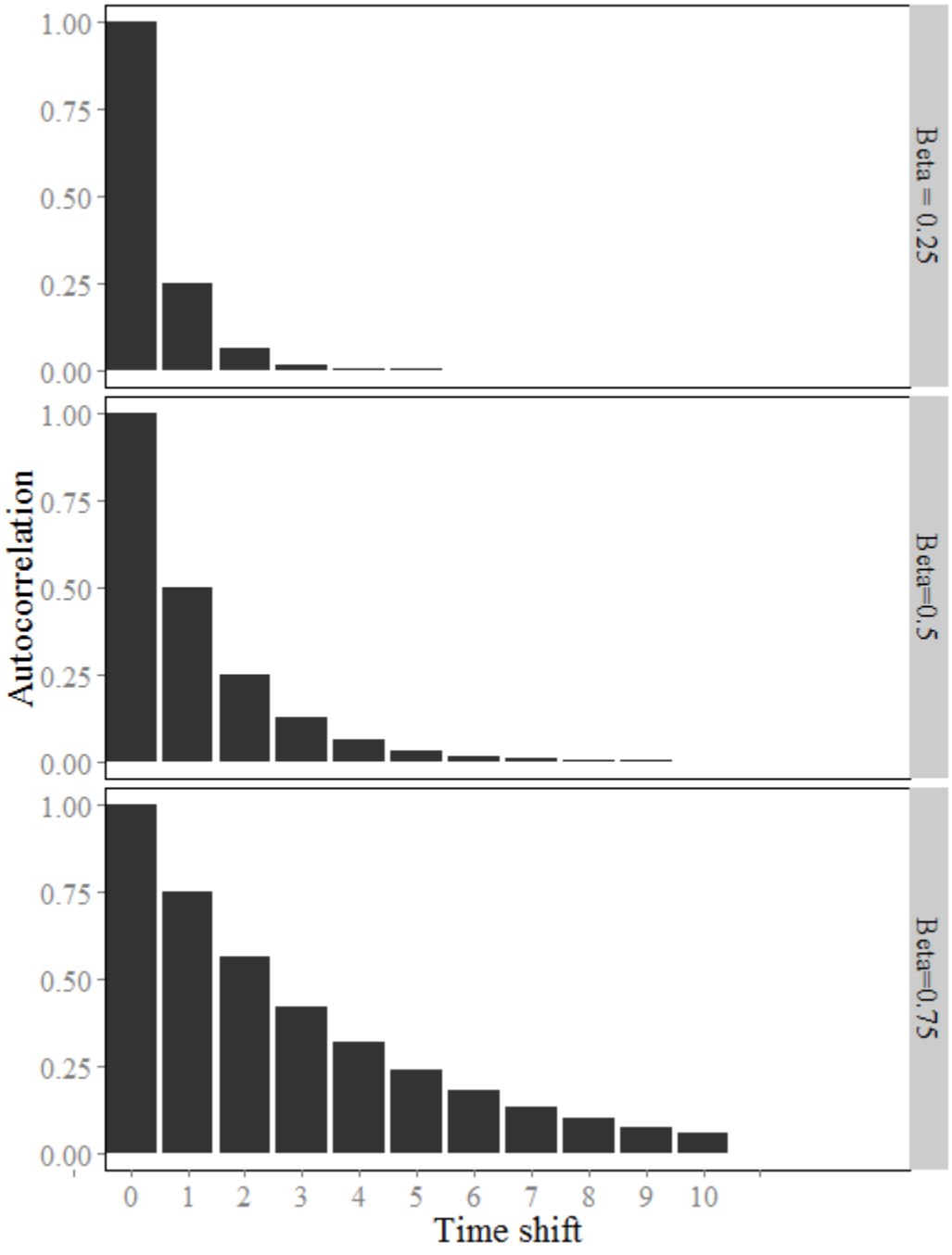
$$= c + \sum_{i=0}^{T} \varphi^i \varepsilon_{t-i} + \sum_{i=0}^{T} \varphi^i \varepsilon_{t-i} \qquad (15)$$

with all the above terms identical to Equation 14 above. Consider the limit as the term T becomes arbitrarily large. Supposing that φ is greater than one in absolute value, the term y_t will also increase in absolute value as the process continues, meaning that the average of the observed data would depend on the time point, violating weak stationarity. Suppose instead that φ is less than one in absolute value. Then, y_t may be expressed as

$$y_t = \frac{c}{1-\varphi} + \sum_{i=0}^{T} \varphi^i \varepsilon_{t-i} \qquad (16)$$

and the long-term expected value of the y_ts is $\frac{c}{1-\varphi}$, since the expected value of the error terms is assumed to be zero. Moreover, one may show analytically (e.g., Hamilton, 1994, p. 53) that the variances and covariances of this time series process remain constant across time points, satisfying weak stationarity. Specifically, the j^{th} autocorrelation is equal to φ^j, meaning that then autocorrelation function decays geometrically over time. The autocorrelation function given different values of φ is plotted in Figure 5, below.

Figure 5. The autocorrelation function of an AR(1) process, given different values of the autoregressive parameter.



As indicated in Figure 5, in an AR(1) process, every observation is correlated with every other observation. However, depending on the value of the

autoregressive parameter, this correlation may be quite small for data points that are far apart from each other in time. This sort of time series structure could be applicable to ability measurements in educational psychology, in which an ability measurement at a particular time point gives information about ability at a later time (see, e.g., Jordan, Kaplan, & Hanich, 2002, Catts, Bridges, Little, & Tomblin, 2008).

However, recall that the assumption of weak stationarity given the model must hold for such an autoregressive model to be employed. In particular, the assumption that the observed means remain the same over time given the model may be untenable without a growth term, particularly when student growth is the construct under study. It is therefore necessary to include some sort of growth term in an autoregressive model to properly apply it to educational psychology. In adding this growth term, a second-order autoregressive latent growth model will be outlined.

To introduce a growth term to the autoregressive unidimensional time series model, let the vector $\boldsymbol{\eta_t}$ in Equation 13 equal the column vector with elements 1, t, and θ_t. Then specify the matrix $\boldsymbol{\Gamma}$ as

$$\boldsymbol{\Gamma} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ \alpha & \beta & \varphi \end{bmatrix} \qquad (17)$$

with intercept term α, growth term β, and autoregressive parameter φ. Then, following Equation 13, $\boldsymbol{\eta_{t+1}}$ becomes:

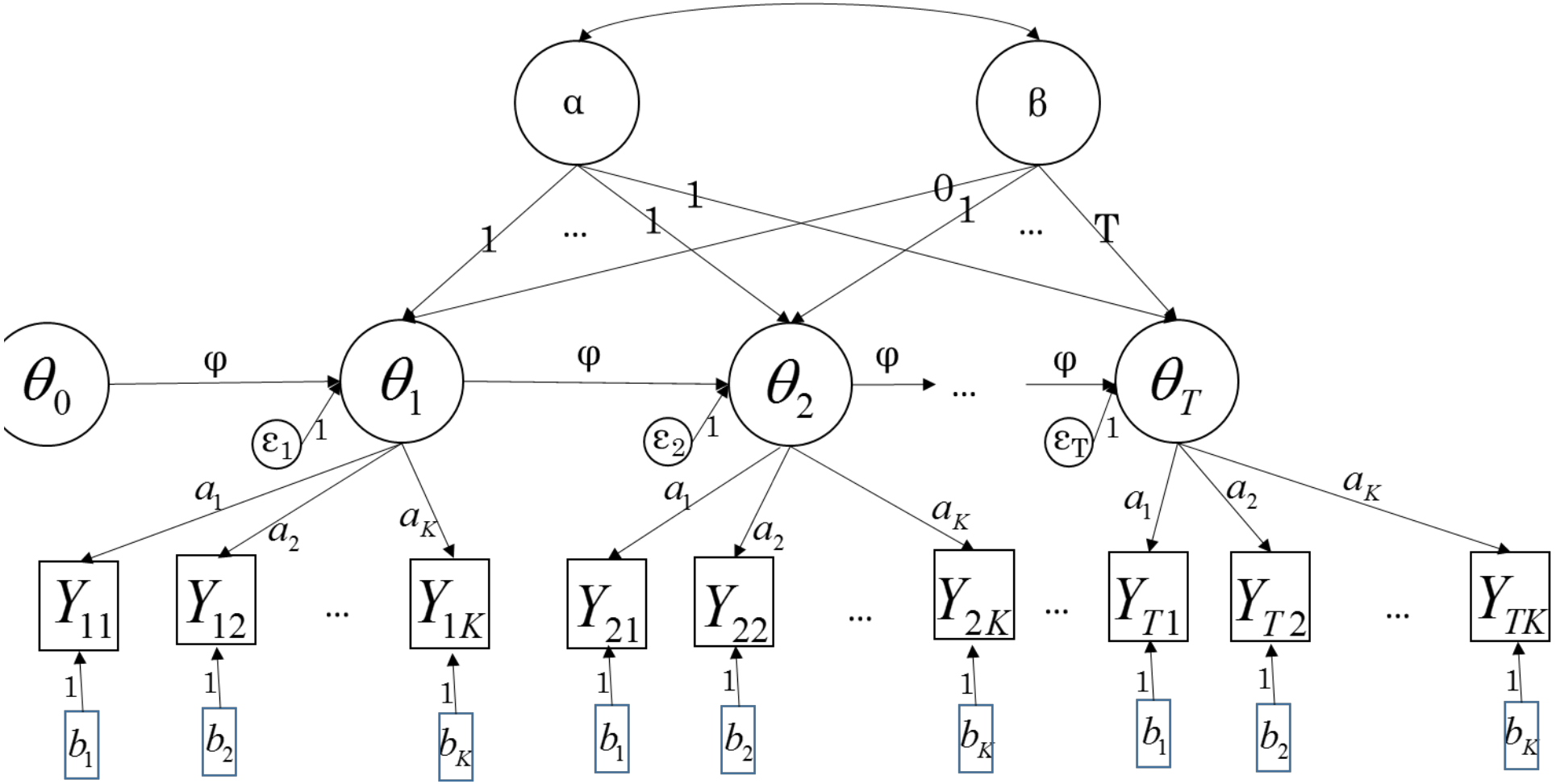
$$\begin{bmatrix} 1 \\ t+1 \\ \theta_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & t \\ \alpha & \beta & \varphi & \theta_t \end{bmatrix} + \boldsymbol{\zeta_t} \qquad (18)$$

with the first element of the error vector $\boldsymbol{\zeta_t}$ set equal to zero. This specification is similar to the specification in Equation 14, however the first element of $\boldsymbol{\eta_t}$ may be used to generate a random intercept term in Equation 12, while the second element of $\boldsymbol{\eta_t}$ may be used to generate a random slope term in Equation 12. In other words, including the constants 1 and t in $\boldsymbol{\eta_t}$ allows for the estimation of individual-level growth and intercept terms. To allow these growth and intercept terms to vary, call the column vector consisting of the elements α, β, and φ $\boldsymbol{\xi}$. Finally, let $\boldsymbol{\xi}$ be distributed as a multivariate Normal distribution, with notation similar to Equation 8:

$$\boldsymbol{\xi} = \boldsymbol{\mu_\eta} + \boldsymbol{\zeta_\eta} \qquad (19)$$

with mean vector $\boldsymbol{\mu_\eta}$ and disturbance vector $\boldsymbol{\zeta_\eta}$; call the variance-covariance matrix of this multivariate Normal distribution $\boldsymbol{\Psi}$, similarly to Equation 3, above. A second-order autoregressive latent growth model with this specification is shown in Figure 6, below

gure 6. A unidimensional second-order autoregressive latent growth model.



This second-order autoregressive latent growth model may be further modified or altered in many of the ways previously described in the latent growth modeling sections. For instance, researchers may be interested in including explanatory variables which may account for the variance in the intercept and growth terms of the model. Also, depending on the structure of the data, there may be many observed items collected at any particular time point, which may or may not be dichotomous. Again, supposing that the data are dichotomous, the link function shown in Equation 9 may be employed. In this way, all the advantages of second-order latent growth modeling may be retained while also including an autoregressive structure. Two measurements of student ability may remain significantly correlated even after correcting for that student's initial ability and growth, which would make the inclusion of an autoregressive term necessary. As discussed later, simulation studies indicate that failing to model an autoregressive trend when the data are generated with an autoregressive trend inflates type 1 error rates and biases and reduces power.

In simulation studies that model autoregressive processes, it is common to also model a moving average trend, another time series process (Hamilton, 1994). A moving average trend is defined by allowing previous error terms to have an impact on current observations:

$$y_t = c + \varepsilon_t + \theta\varepsilon_{t-1} \qquad (20)$$

where θ is the moving average parameter and y_t, c, ε_t, and ε_{t-1} are defined as in Equation 14. It may be shown analytically (Hamilton, 1994) that the first-order autocorrelation, that is, the correlation between the observations of two adjacent time points, is equal to $[\theta / (1 + \theta^2)]$. On the other hand, any higher-order autocorrelations are assumed equal to zero (Hamilton, 1994). In other words, the moving average time series structure allows observations that are adjacent in time to be correlated, while not forcing every point to be correlated with every other point.

However, as noted above, in educational psychology there are theoretical reasons to expect every observation to correlate with every other observation. More commonly, autoregressive and moving average trends are both modeled at the same time, in an autoregressive-moving average (ARMA) process:

$$y_t = c + \varphi y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1} \qquad (21)$$

with y_t, y_{t-1}, c, φ, ε_t, θ, and ε_{t-1} as defined in Equations 19 and 20. The specification in Equation 21 is called an ARMA(1, 1) process, since there is one autoregressive parameter and one moving average parameter. The autocorrelation function for an ARMA(1, 1) process is more complex mathematically (Hamilton, 1994), but it may be thought of as being similar to the autocorrelation function for an autoregressive process, with the first-order autocorrelation being altered by the moving average parameter.

When a time series trend is not adequately modeled by these autoregressive or moving average processes, researchers may instead decide to freely estimate each

autocorrelation, rather than estimating them according to the autocorrelation functions provided above. This approach involves estimating more parameters than any of the previously discussed time series models. That is, for t time points, this would involve estimating $\frac{t(t-1)}{2}$ autocovariances (Kwok, West, & Green 2007). In contrast, the autocovariances for a first-order autoregressive process and a first-order moving-average process are completely specified by estimating a single parameter (Kwok, West, & Green, 2007). Moreover, the autocovariances of an ARMA(1, 1) process may be specified by estimating only two parameters, the θ and φ parameters.

There are a number of possible errors one might make when attempting to model a time series process. For instance, one might freely estimate all the autocovariances when an autoregressive trend would adequately model the data. Such a model would be over-parameterized. On the other hand, one might model all the observations as being independent when an autoregressive trend is actually present. This model would be under-parameterized. Finally, one might model an autoregressive trend as a moving average trend, or vice-versa, which would be generally mis-specified. Simulation studies have evaluated the practical impact that these errors have on estimation.

Simulation studies have indicated that failing to model an autoregressive term when such a trend actually exists may bias the estimation of a number of relevant parameters for latent growth modeling. In a simulation study, Murphy and Pituch (2009) assessed the parameter recovery of a first-order latent growth model when time series trends were present. Failing to model these time series trends

resulted in an elevated type 1 error rate when testing fixed effects. That is, for conditions with smaller samples, (N = 30), failing to model a time series trend resulted in type 1 error rates ranging from 0.079 to 0.125, when the nominal type 1 error rate was 0.05. This means that failing to include a time series component may lead researchers to incorrectly state that a variable explains student growth, or that a variable explains the students' initial ability.

More relevantly, Kwok, West, and Green (2007) simulated a condition in which data were generated with an autoregressive trend, but estimated without a time series trend. That is, longitudinal data were generated so that previous constructs impacted future constructs as in Equation 14, but the constructs were estimated as second-order latent growth models without including an autoregressive parameter. Then, relative bias of standard error and variance terms were presented, along with estimation of statistical power. When the estimation model was underspecified compared to the way the data were generated, the power to detect a nonzero intercept or slope was 0.28, but when the estimation model was correctly specified, this power term was 0.32. In fact, the over-specified estimation models, on average, had the greatest power to detect a nonzero effect, with a power of 0.38. This suggests that even if a time series trend is not present in the population, including a time series trend may improve power. On the other hand, the estimation model had no impact on type 1 error rates for testing the intercept and slope terms of a latent growth model.

Following these analyses, Murphy, Beretvas, and Pituch (2011) incorporated autoregressive and ARMA(1,1) processes to generate and estimate second-order

latent growth models. Varying the sample size, the length of the time series, and the generation and estimation methods, the effects of model mis-specification on parameter estimation were quantified as the mean relative bias term, computed as

$$RPB(\hat{\theta}) = \frac{\bar{\hat{\theta}} - \theta}{\theta} \qquad (20)$$

where θ is the true parameter value and $\bar{\hat{\theta}}$ is the mean parameter estimate. According to Hoogland and Boomsma's (1998) work, a mean relative bias term greater than 0.05 in absolute value can be considered substantially biased.

When the data were generated using an autoregressive process similar to Equation 14, the mean relative bias of the slope and intercept terms in the latent growth models had less than 0.05 mean relative bias. However, the data generation method had an impact on the estimation of the variance of the slope and intercept terms. That is, when a large autocorrelation was used to generate the data (with φ = 0.8), estimating the data without an autoregressive term resulted in severe underestimation (of degree 92.1% up to 94.1%) of the variance of the intercept. On the other hand, correctly estimating the data with an autoregressive term resulted in mean relative biases between -0.332 and -0.719. Although handling autoregressive data using a model that correctly specified the autoregressive structure did improve the mean relative bias of the variance of the intercept term, these estimates were still regarded as being substantially biased, according to Hoogland and Boomsma's (1998) work.

Estimating the variance of the slope was also substantially biased when a large autoregressive term was used to generate the data. Again, the bias was reduced when the estimation correctly included an autoregressive term. Specifically, when the estimation did not include an autoregressive term, the mean relative bias of the slope variance was between -0.616 and -0.703, whereas when the estimation did include an autoregressive term, mean relative bias was between -0.044 and -0.300.

Finally, the estimated covariance between the intercept and slope was substantially biased when an autoregressive structure was simulated, regardless of the size of the generating autoregressive term or the estimation method. When the model was estimated without an autoregressive term, mean relative bias ranged between -0.155 and -0.993, depending mostly on the size of the autoregressive term. However, when an autoregressive estimation model was estimated, the absolute mean relative bias ranged between 0.027 and 1.972, depending mostly on the sample size of the observations.

To summarize, estimating an autoregressive term when an autoregressive term was present resulted in somewhat less biased variance terms for the intercept and slope in Murphy et al.'s (2011) work. However, even when the autoregressive structure was correctly specified, estimates of these terms were substantially biased. Moreover, when the generating autocorrelation was low ($\varphi = 0.3$), estimation of between 14.3% and 43.7% of the models which included an autoregressive parameter failed to converge, depending on the sample size. Also, when the

generating autocorrelation was high ($\varphi = 0.8$), estimation of between 1.3 and 12.7% of the autoregressive models failed to converge, again depending on the sample size.

This proportion of non-convergent cases, along with the substantially biased variance terms, should lead researchers to consider alternative estimation methods. Murphy et al. (2011) used MPlus to estimate these second-order autoregressive latent growth models, which employs likelihood-based estimation, relying on iteratively updating proposed model estimates to check a discrepancy function (Muthén, & Muthén, 2007). These discrepancy terms may achieve a local minimum; furthermore, there is no theoretical guarantee that the estimation method will result in a converged solution.

In contrast, Bayesian methods, and in particular the Gibbs sampler, are theoretically guaranteed to converge eventually (Gelman, Carlin, Stern, & Rubin, 2014). In Gibbs sampling, the conditional distribution of each parameter given all the other parameters is obtained. Then, parameters are initialized at particular values. Next, parameters are sampled according to the conditional distribution, plugging in the most recent sample values for each of the other parameters. This process repetitively continues until, eventually, distributions for each parameter are obtained. Often, the procedure starts with some number of independent samples of values, called chains, and samples are taken from each of the chains separately (Brooks & Gelman, 1998). Then, the variance in the samples between and within chains is compared to estimate a statistic called the scale reduction factor which provides evidence that the chains have converged (Brooks & Gelman, 1998). Once the chains have converged, one may obtain point estimates of parameters, for

instance from the median value of the simulated posterior distribution, or obtain credible intervals, which are somewhat similar to confidence intervals in frequentist statistics (Gelman et al., 2007).

Gibbs samplers for estimating models similar to a first-order latent growth model already exist in the educational psychology literature. For instance, Fox and Glas (2001) outline a Gibbs sampler for a multilevel IRT model. In that work, the researchers were contextualizing the estimation process as a means of estimating an item response theory model for students nested within schools. However, one might use a quite similar algorithm to estimate an item response theory model for measurement occasions nested within examinees. Then, by including a growth term and an intercept for each student, one might apply Fox and Glas's (2001) work to estimate a model similar to Figure 3. However, Fox and Glas (2001) do not explicitly attempt to estimate a model that has an autoregressive component, in fact, it is assumed that observations do not correlate with each other, given the model. As previously discussed, if an autoregressive trend is present in the population and it remains unmodeled, simulation studies indicate that power may be reduced and the actual type 1 error may be inflated for testing growth terms. To account for the autocovariance structure implied by an autocorrelation term, the Kalman filter is introduced.

Bayesian Methods: The Kalman Filter

Introduced by Kalman (1960), the Kalman Filter is a means of sampling latent variables when they are related in an autocorrelated manner. Conceptually, the Kalman Filter proceeds in four steps, the first three of which may be considered

together as the forward filter. In the forward filter, first the prior for the latent variable θ_t is obtained, given the previous value of that latent variable, θ_{t-1}. Next, the likelihood for the observed data y_t is obtained, once again given θ_{t-1}. Finally, with the prior and the likelihood for θ_t, the posterior distribution for θ_t is obtained, given the data y_t. The forward filter proceeds in this manner until posterior distributions are obtained for latent variables θ_1, θ_2, …, θ_T. The final step of the Kalman filter, backwards sampling, updates the value for each latent variable θ_t given the next value of that latent variable, θ_{t+1}. The exact mathematics of the Kalman filter will now be delineated; similar derivations may be found in Hamilton (1994), or also in Petris, Petrone, and Campagnoli, (2009).

Consider the following time series structure:

$$y_t = \theta_t + \varepsilon_t \qquad (21)$$

$$\theta_t = \alpha + \varphi * \theta_{t-1} + \zeta_t \qquad (22)$$

with observed variable y_t, latent variable θ_t, error vectors ε_t and ζ_t, latent intercept α, and autoregressive parameter φ. Note that this is a particular application of Equations 12 and 13. For the sake of stationarity, it is assumed that φ is less than one in absolute value, and it is assumed that the error terms ε_t and ζ_t are uncorrelated at each time point. Also, it is assumed that ε_t and ζ_t are Normally distributed, with means of zero and variances of σ^2 and τ^2, respectively. Finally, let D_t be the set containing elements $y_1, y_2, \ldots, y_t$.

In a Gibbs sampler, the conditional posterior distribution of θ_t is sought, given all the other parameters. Thus, it is assumed that α, φ, σ^2, and τ^2 are all

known. As mentioned above, first the prior for θ_t is sought, given the posterior distribution of θ_{t-1}. Suppose for the sake of mathematical induction that the posterior for θ_{t-1} given D_{t-1} is Normally distributed, with a mean of m_t and a variance of C_t. Then, the prior for θ_t given D_{t-1} is given by

$$p(\theta_t \mid D_{t-1}) = \int p(\theta_t \mid \theta_{t-1}, D_{t-1}) * p(\theta_{t-1} \mid D_{t-1}) d\theta_{t-1} \qquad (23)$$

for the previous data D_{t-1}, previous latent estimate θ_{t-1}, and current latent estimate θ_t. Now, by assumption, the posterior distribution $p(\theta_{t-1} \mid D_{t-1})$ is Normally distributed, with a mean of m_{t-1} and a variance of C_{t-1}. Note also that $p(\theta_t \mid \theta_{t-1}, D_{t-1})$ is given by Equation 22, that is, θ_t is Normally distributed, with a mean of $\alpha+\varphi*\theta_{t-1}$ and a variance of τ^2. This integration, then, is analogous to computing the posterior distribution of a variable with a Normal prior, given by $p(\theta_{t-1} \mid D_{t-1})$, and a Normal likelihood, given by $p(\theta_t \mid \theta_{t-1}, D_{t-1})$. A common result in Bayesian statistics (e.g., Gelman, Carlin, Stern, & Rubin, 2007) is that, after computing this integration, $p(\theta_t \mid D_{t-1})$ is Normally distributed, with a mean of $\alpha+\varphi m_{t-1}$ and a variance equal to the quantity $\varphi^2 C_{t-1}+\tau^2$; call these quantities a_t and R_t, respectively.

In the next step of the Kalman filter, the predictive distribution for y_t, given D_{t-1} is obtained in a similar fashion:

$$p(y_t \mid D_{t-1}) = \int p(y_t \mid \theta_t, D_{t-1}) * p(\theta_t \mid D_{t-1}) d\theta_t \qquad (24)$$

with $p(y_t \mid D_{t-1})$ the predictive distribution for y_t given the previous data D_{t-1}. Note that the second term of the integration, $p(\theta_t \mid D_{t-1})$, was just shown to be a Normal

distribution, with a mean of a_t and a variance of R_t. Next, recall that $p(y_t \mid \theta_t, D_{t-1})$ is given by Equation 21, where y_t is Normally distributed, with a mean value equal to the expected value of θ_t, given D_{t-1}, and a variance equal to σ^2. Above, it was shown that the expected value of θ_t is equal to m_t. Thus, the term in the integration is equal to two Normal distributions, each with the same mean, one with a variance of σ^2 and one with a variance of R_t. The result of such an integration (shown, e.g., in Gelman et al., 2007) is a Normal distribution, with a mean equal to a_t and a variance equal to σ^2+R_t; call this variance term Q_t.

To obtain the posterior distribution for θ_t given D_t, consider the bivariate normal distribution of θ_t and y_t, given D_{t-1}. Once this bivariate Normal distribution is obtained, it will be possible to express θ_t given D_t. Since the univariate Normal distributions of θ_t and y_t given D_{t-1} were obtained above, it only remains to compute the covariance of y_t and θ_t, given D_{t-1}. Recall that this covariance term is defined as

$$Cov(y_t, \theta_t) = E[(y_t - E\{y_t\}) * (\theta_t - E\{\theta_t\})] \qquad (25)$$

where $E(x)$ denotes the expected value of the variable x. Now, $E(y_t) = E(\theta_t)$, as described above, and, by Equation 21, $y_t = \theta_t + \varepsilon_t$. In that case, Equation 25 may be rewritten as

$$Cov(y_t, \theta_t) = E[(\theta_t + \varepsilon_t - E\{\theta_t\}) * (\theta_t - E\{\theta_t\})]$$

$$= Var\{\theta_t\} + E[\varepsilon_t * (\theta_t - E\{\theta_t\})] \qquad (26)$$

where the second term may be rewritten as

$$E[\varepsilon_t * (\theta_t - E\{\theta_t\})] = E[\varepsilon_t * \xi_t] \qquad (27)$$

and, by assumption, ε_t and ξ_t are independent, meaning that the expectation of their product is equal to the product of their expected values, which are both equal to zero. In short, the covariance of y_t and θ_t is equal to the variance of θ_t, which was called R_t. In that case, the bivariate normal distribution of y_t and θ_t given D_{t-1} is a bivariate Normal distribution with mean vector $[a_t, a_t]^T$ and variance-covariance matrix $\begin{bmatrix} R_t & R_t \\ R_t & Q_t \end{bmatrix}$. Following standard results on the bivariate Normal distribution, (see. e.g., Jensen, 2000), the posterior distribution of θ_t given D_t is itself Normally distributed, with a mean m_t equal to $a_t + \frac{R_t}{Q_t}$ and a variance C_t equal to $R_t - R_t * \frac{R_t}{Q_t}$ .In short, if the Kalman filter is initialized with a variable θ_0 with a particular fixed mean m_0 and a particular variance C_0, the forward filter is able to obtain closed-form solutions for the conditional distribution of each latent variable θ_t.

However, in an autoregressive structure, a latent variable θ_t also depends on the next data point y_{t+1}. In the backwards sampling step of the Kalman Filter, this dependency is exploited and posterior distributions for θ_t are updated based on y_{t+1}. To see that the value θ_t depends on y_{t+1}, consider the bivariate Normal distribution of θ_t and y_{t+1} given D_t. As discussed above, the mean and variance of θ_t given D_t is m_t and C_t, respectively. Also, the mean and variance of y_{t+1} given D_t is a_{t+1} and Q_{t+1}, since the distribution of θ_{t+1} given D_t is exactly the same as the that distribution sought in the second step of the Kalman filter, above. To fully specify this bivariate Normal distribution, consider the covariance between θ_t and y_{t+1}. By the properties of covariance, and by Equation 22,

$$Cov(y_{t+1}, \theta_t) = Cov(\alpha + \varphi\theta_t + \zeta_{t+1} + \varepsilon_t, \theta_t)]$$

$$= Cov(\varphi\theta_t, \theta_t) = \varphi Var(\theta_t) \qquad (26)$$

with the next-to-last equality following from the assumption that the error terms at each time point are uncorrelated with each other. Then, the bivariate Normal distribution of θ_t and y_{t+1} has a mean vector of $[m_t, a_{t+1}]$ and a variance-covariance matrix equal to $\begin{bmatrix} C_t & \varphi C_t \\ \varphi C_t & Q_t \end{bmatrix}$. Again using standard results about the bivariate Normal distribution, the mean of θ_t given D_{t+1} is given by $m_t + \varphi\frac{C_t}{Q_t}(y_t - a_t)$ and its variance is equal to $C_t - \varphi^2\frac{C_t^2}{Q_t}$. Recall that without this backward sampling step, the variance was equal to C_t, meaning that the subtracted term is variance in the latent estimate that is accounted for by this backward sampling step.

Multivariate extensions of the Kalman filter, along with adaptations for also including moving average processes are available in the literature (e.g., Hamilton, 1994). Taken together, the Kalman filter and the work by Fox and Glas (2001) may be used to construct a Gibbs sampler for a second-order autoregressive latent growth model. However, the work by Fox and Glas (2001) focuses mainly on the recovery of item parameters, while the parameters of interest in latent growth modeling are the latent intercept and growth terms. Moreover, that work did not incorporate an autoregressive structure.

Statement of Purpose

Murphy et al. (2013) investigated the estimation of a second-order autoregressive latent growth model, with results indicating generally underestimated variance terms and high rates of non-convergence even when the correct model was estimated. The proposed estimation method may further improve upon this analytic method. The proposed method is Bayesian in nature, and is therefore theoretically guaranteed to converge eventually. The method incorporates the Kalman filter, a method specialized for time series data.

While similar to previous work (Fox & Glas, 2001), the proposed model differs from Fox and Glas's (2001) analysis in one major way. Recall that Fox and Glas (2001) proposed a multilevel IRT model using Gibbs sampling. The proposed model will use measurement occasions at the first level of the multilevel model, and person-wide ability estimates at the second level. However, unlike Fox and Glas's (2001) analysis, this proposal includes an autoregressive component, which converges more efficiently using the forward filtering backwards sampling algorithm.

Once the Gibbs sampler is outlined, data will be simulated and estimated with and without the Kalman filter. First, all the convergence diagnostics will be presented. Next, for the level-one and level-two variance terms, and the level-two means, point estimates for parameters, taken as median values of the chains, will be presented, along with Bayesian credible intervals. Finally, median item parameters, level-one regression weights, and latent ability estimates will be graphed against their true values, and correlations will be reported.

There are three research questions that this work aims to address. Firstly, does using the Kalman filter result in the convergence of more estimated terms? Secondly, how similar are the estimated terms to the true values, with and without using the Kalman filter?

Methods

Data Generation: Latent Growth Model Sample

A dataset will be simulated and analyzed to assess the bias of parameters estimated using the proposed estimation method for a second-order autoregressive Bayesian latent growth model.

Most of the generating parameters in this analysis are specified in Equation 18, reprinted here for convenience:

$$\begin{bmatrix} 1 \\ t+1 \\ \theta_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & t \\ \alpha & \beta & \varphi & \theta_t \end{bmatrix} * + \boldsymbol{\zeta_t} \qquad (29)$$

with the time parameter t coded as zero for the first measurement occasion, one for the second measurement occasion, and so on. Persons' ability levels at the first time point, or a in the above, will be generated in two steps. First, person-level ability estimates will be taken as random draws from the standard normal distribution. This will simulate differences in initial ability. Next, occasion-specific error for the first time point will be simulated by setting a equal to a draw from the Normal distribution, with a mean at the person-level ability estimate, and variance equal to one-third, corresponding to a medium level of occasion-specific variance (from Geiser, Keller, & Lockhart, 2013). That is, the total variance for the first time point will be equal to four-thirds, with a variance of one associated with true variance in ability and a variance of one-third associated with time-specific measurement error. To match Geiser, Keller, and Lockhart's (2013) work, 1,000 simulees will be generated.

Next, the autoregressive parameter (φ) in the above will be taken as a draw from a truncated normal distribution, with a mean of 0.65 and a variance of 0.273; this approximates the range between 0.5 and 0.8 which was investigated in Price, (2012) for a model analogous to a first-order autoregressive latent growth model.

The growth parameter (ß) in the above Equation will be generated as a draw from a normal distribution, with a mean of 0.34, and a variance of 0.17, corresponding to Rosenbaum's (1986) estimate of high school student growth in mathematics, per semester. Using these values of α, β, φ, and θ_1, future time-specific ability estimations will be generated following Equation 18. Matching previous research in second-order latent growth modeling, four time points will be simulated (Geiser, Keller, & Lockhart, 2013).

Finally, once values for θ_t are generated, individual items for specific persons at specific time points will be generated using the following formula:

$$P(\mathbf{y}=1) = \Phi(\mathbf{b} + \mathbf{a}*\boldsymbol{\theta}_t) \qquad (30)$$

with $\Phi(x)$ representing the cumulative density function of the standard normal distribution, $\mathbf{b}$ representing the vector of difficulty parameters, and $\mathbf{a}$ representing the vector of item discrimination parameters. Values for the 30 items' discrimination and difficulty parameters will match those reported for a previously calibrated ACT mathematics test, and will be used to generate binary responses at each time point. In other words, the item parameters will be simulated as being constant over time, with no item drift or any substitutions of different items at different times. For identification purposes, the first elements of $\mathbf{b}$ and $\mathbf{a}$ will be set

to zero and one, respectively, for both generation and estimation of these data. These item parameters are reported in Table 1, below.

Table 1. Proposed item parameters for simulation study.

Item number	Discrimination parameter *a*	Difficulty parameter *B*
1	1	0
2	0.847535	-1.00585
3	0.701955	-1.22521
4	0.902016	-1.36332
5	1.046108	-0.28199
6	0.890067	-0.57313
7	0.893602	-1.36773
8	0.716244	-0.85472
9	1.164358	-0.4441
10	0.368682	-0.32641
11	0.591835	0.074493
12	0.965367	0.030348
13	0.829711	0.181143
14	1.442186	0.212373
15	1.039885	-0.28632
16	0.905395	-0.04697
17	1.45784	-0.01277
18	1.222484	0.171493
19	0.867419	0.204116
20	1.006531	0.324807
21	0.948628	0.345869
22	0.855386	1.452706
23	0.999414	0.731968
24	1.136003	1.229729
25	0.804661	1.273888
26	0.967754	1.803751
27	1.240903	1.047733
28	1.589016	1.626541
29	1.446432	1.895972
30	1.364186	2.806588

Data Generation and Estimation: Item Calibration

Simulating initial item calibration will proceed in three steps. First, ability levels θ_1 will be generated from a Standard Normal distribution. Then, responses will be generated following Equation 30, with values for **a** and **b** equal to those used in generating the Latent Growth Model simulated dataset. In other words, it will be assumed that items are calibrated at one specific time point, and that the item parameters do not change over time. Finally, item parameters will be estimated using the package "mirt" in R (Chalmers, 2012).

Outline of the Gibbs Sampler

After the data are simulated, they will be estimated twice, once using the Kalman filter, and once without using the Kalman filter. To run this Gibbs Sampler, the software packages R and Rcpp will be used. As previously stated, the Gibbs Sampler will largely be taken from Fox and Glas's (2001) work, with only two major adjustments. Mainly, rather than the first level corresponding to students and the second level corresponding to schools, in this analysis the first level corresponds to measurement occasions and the second level corresponds to examinees. The measurement model is given by

$$P(\mathbf{y}=1) = \Phi(\mathbf{b} + \mathbf{a}*\boldsymbol{\theta}_t) \qquad (31)$$

with notation identical to Equation 29. Then, the first level of the model is given by

$$\begin{bmatrix} 1 \\ t+1 \\ \theta_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & * \ t \\ \alpha & \beta & \varphi & \theta_t \end{bmatrix} + \boldsymbol{\zeta}_t \qquad (32)$$

with notation the same as Equation 30. Call the last element of $\boldsymbol{\zeta}_t$ e_t; then, e_t is assumed to be Normally distributed, with a mean of zero and a variance called σ^2. Finally, in the second level, the terms α, β, and φ are assumed to follow a multivariate Normal distribution, with means of $\lambda_{\alpha 0}$, $\lambda_{\beta 0}$, and $\lambda_{\varphi 0}$, respectively, and covariance matrix $\mathbf{T}$; call the vector containing the terms α, β, and φ, $\boldsymbol{\beta}_i$ for a given individual i.

Before detailing the algorithm discussed in Fox and Glas (2001), the adjustments made to the algorithm are mentioned first. The adjustments are twofold. Firstly, for one estimation, the latent constructs θ will be sampled using the filter-forward backward sampling algorithm, that is, the Kalman filter as

described in the literature review above. Secondly, the autoregressive parameter φ will be estimated separately from the other level-two regression weights, since, for stationarity purposes, φ must be constrained to be less than one in absolute value. To accomplish this, φ will be estimated assuming knowledge of every other parameter and sampled as a truncated Normal distribution, with truncation points at negative and positive one. The mean and variance of this truncated Normal distribution will be computed identically to how these terms are computed for the other level-two regression weights, described below.

Following Fox and Glas, (2001), in the first step latent variables $\mathbf{Z}$ are sampled using truncated Normal distributions, each with means of $\mathbf{b} + \mathbf{a}*\boldsymbol{\theta}_t$ and variances of one. To reflect the binary nature of the response data, the latent variables $\mathbf{Z}$ are constrained to be greater than zero if the corresponding question was correctly answered and are constrained to be less than zero otherwise.

In the next step, estimates $\boldsymbol{\theta}_t$ will be sampled using the Kalman filter, as described above. For more details, see the section on this topic in the literature review, above. When the $\boldsymbol{\theta}_t$ estimates are taken without using the Kalman filter, the estimation procedure will follow the procedure outlined in Fox and Glas (2001). Define quantities $\underline{v}$ and $\hat{\theta}_{ti}$, where

$$v = \frac{1}{\sum_{k=1}^{K} a_k^{\,2}} \qquad (33)$$

and

$$\hat{\theta}_{ti} = v * \sum_{k=1}^{K} a_k (Z_{tik} + b_k) \qquad (34)$$

respectively, with a_k, and b_k, referring to item discrimination and difficulty parameters, respectively. Then each estimate θ_{it} will be taken as a draw from a Normal distribution, with a mean equal to $\frac{\hat{\theta}_{it}/\nu + \mathbf{x}*\beta/\sigma^2}{1/\nu + 1/\sigma^2}$ and a variance equal to $\frac{1}{1/\nu + 1/\sigma^2}$, where $\mathbf{x}$ refers to the vector $[1\ t\ \theta_t]^t$ and β refers to that person's ability, growth, and autoregressive regression weights. For more information on the derivation of these quantities, see Fox and Glas (2001).

In the third step, the item parameters in the $\mathbf{a}$ and $\mathbf{b}$ vectors will be sampled. These parameters will be taken as a draw from the multivariate Normal distribution. Let $\boldsymbol{\xi}_\mathbf{k}$ be the vector containing item parameters a_k and b_k, respectively, and let $\mathbf{Z}_\mathbf{k}$ be the vector containing all those elements of $\mathbf{Z}$ corresponding to item number k. Then, as per Fox and Glas (2001),

$$\mathbf{Z}_\mathbf{k} = [\ \boldsymbol{\theta}\ \mathbf{-1}\ ]\ \boldsymbol{\xi}_\mathbf{k} + \boldsymbol{\varepsilon}_\mathbf{k} \qquad (33)$$

where $\boldsymbol{\varepsilon}_\mathbf{k}$ is a Normally distributed random variable with a mean of zero and a variance of one. The vector $\boldsymbol{\xi}_\mathbf{k}$ may then be sampled as a multivariate Normal, with mean $\hat{\xi}_k$ corresponding to Ordinary Least Squares regression weights and covariance matrix $([\ \boldsymbol{\theta}\ \mathbf{-1}\ ]^{T} * [\ \boldsymbol{\theta}\ \mathbf{-1}\ ])^{-1}$.

After the item parameters are sampled, the terms α, and β are sampled, followed by separately sampling φ, as described above. In the former step, α and β are sampled again according to Fox and Glas (2001). Define $\mathbf{d} = \sigma^2 * \boldsymbol{\theta} + \mathbf{T}_\mathbf{a}^{-1}\boldsymbol{\gamma}$, where $\boldsymbol{\gamma}$ is the vector containing γ_α and γ_β, and $\mathbf{T}_\mathbf{a}$ is the principal submatrix of $\mathbf{T}$ formed by omitting the final row and column. Next define $\mathbf{D} = (1/\sigma^2 + \mathbf{T}_\mathbf{a}^{-1})^{-1}$. Then, the

terms α and ß will be sampled from a multivariate Normal distribution with mean $\mathbf{Dd}$ and covariance matrix $\mathbf{D}$.

Next all the elements of $\boldsymbol{\gamma}$ are sampled, continuing to mirror the algorithm described in Fox and Glas (2001). In this application, define $\mathbf{E} = \mathbf{T} * 1/n$, and define $\mathbf{e} = \sum_{i=1}^{N} \boldsymbol{T}^{-1} * \boldsymbol{\beta_i}$. Then, $\boldsymbol{\gamma}$ is drawn as a multivariate Normal distribution, with mean $\mathbf{Ee}$ and covariance matrix $\mathbf{E}$.

In the second-to last step, the error variance term σ^2 is sampled, again following Fox and Glas (2001). Let $\mathbf{X}_t$ equal [1, t, θ_t], for a particular time point *t*. Then, let S^2 equal $\frac{1}{n}\sum_{t=1}^{T}(\boldsymbol{\theta_t} - \boldsymbol{X_t\beta})^T(\boldsymbol{\theta_t} - \boldsymbol{X_t\beta})$. The error variance is then taken as an inverse-gamma distribution, with shape parameter N and scale parameter S^2.

Finally, the level-two covariance matrix is sampled in a roughly similar manner to the previous step. Fox and Glas (2001) define a variable $\mathbf{S}$ equal to $\sum_{t=1}^{T}(\boldsymbol{\beta} - \boldsymbol{\gamma})^T(\boldsymbol{\beta} - \boldsymbol{\gamma})$. Then, the covariance matrix $\mathbf{T}$ is sampled as an inverse-Wishart distribution, with t degrees of freedom and scale matrix $\mathbf{S^{-1}}$.

All of these steps together constitute a single iteration of the Gibbs sampler. As described below, these iterations will be repeated some 11,000 times for each time the model is estimated. Finally, three distinct chains will be run for each model to allow for convergence checking using the Gelman-Rubin statistic (Gelman & Rubin, 1994). A parameter will be considered converged only when the Gelman-Rubin statistic is less than 1.2; the number of nonconverged cases will be tallied.

Simulation and Analysis

In this preliminary report, only one dataset will be simulated. For each dataset, following Price (2012), there will be 1,000 burn-in iterations, and 10,000 samples from the posterior distribution. Once the posterior distributions are obtained, mean relative bias for person ability, growth, autoregressive components, item response theory parameters will be computed, with the point estimate taken as the median of the posterior distribution. In the same way, relative bias terms will be computed for the variances and covariances of initial ability, growth, and the autoregressive trend. Mean relative bias estimates, or relative bias estimates, depending on the parameter, will assess parameter recovery. I hypothesize that using the Kalman filter will result in lower mean relative bias, overall. Also, coverage rates for the 95% credible intervals will be computed for these same parameters. Again, it is hypothesized that coverage rates will improve when the Kalman filter is used.

Results

Convergence diagnostics

First, the Gelman-Rubin potential-scale reduction factor (Gelman & Rubin, 1994), also known as R-Hat, will be reported by parameter and condition to assess model convergence. Overall, when using the Kalman filter, most of the model estimates achieved convergence, with the exception of certain ability estimates variance terms. When the Kalman filter was not used, many terms associated with the growth and autoregressive terms failed to converge.

When using the Kalman filter, estimation of a total of fifteen ability estimates failed to converge. Of these, seven were from the second time point, six were from the third time point, and two were from the final time point. Also, the estimation of the level-one variance term failed to converge, with an R-Hat of 1.688. Finally, the level-two variance of the ability estimates failed to converge, with an R-Hat of 1.291.

Without using the Kalman filter, the level-two variance of the growth and autoregressive terms failed to converge. Additionally, the level-two average of the autoregressive term failed to converge. At the first level of the HLM, estimation of 272 individual-level growth terms and of 922 autoregressive terms failed to converge. Finally, estimation of a single item discrimination parameter failed to converge. Notably, the item discrimination parameter had an R-Hat of 1978.38; the final item was quite difficult, and it appears that not many simulees answered it correctly. As a result, there was difficulty estimating the final item discrimination parameter.

Table 2. Convergence Diagnostics (Maximum R-hat) by Estimation Method.

Parameter	Maximum R-Hat	
	Using Kalman Filter	Not Using Kalman Filter
Ability estimates (θ)	1.233	1.056
Item difficulty (b_k)	1.045	1.142
Item discrimination (a_k)	1.176	1978.38
Level-One Regression Weights (β)	1.109	1.479
Level-One Variance (σ^2)	1.688	1.003
Level Two Means (γ)	1.063	4.282
Level Two Variances and Covariances ($\mathbf{T}$)	1.291	1.873

Level-two means, level-one and two variances and covariances

Overall, median estimates and credible interval coverage rates were closer to the true value when the Kalman Filter was used. Kalman filter estimation failed to cover the true initial ability of zero, the variance of the initial abilty, and the variance of the autoregressive parameter. Also, the intervals estimated using the Kalman filter failed to include the true generating ability estimates, item difficulty and discrimination, and the level-one regression weights adequately. See table 4, below, for details.

On the other hand, the credible interval estimates failed to include the true values for the level-two means and variances and the level-one variance, when the Kalman filter was not used. When not using the Kalman filter, every variance term was greatly overestimated. For instance, the level one variance was estimated to be $3.164*10^6$ instead of the true value of one.

ıble 4. Point and Credible Intervals Estimates versus True Values by Estimation Method.

ırameter	True Value	Using Kalman Filter			Not Using Kalman Filter		
		Lower Bound	Posterior Median	Upper Bound	Lower Bound	Posterior Median	Upper Bound
ɛvel-One Variance (σ^2)	0.333	0.045	0.126	0.193	$7.955*10^4$	$3.164*10^6$	$1.077*10^7$
ɛvel Two means (γ)							
Initial ability	0.000	0.366	1.819	2.054	105.59	658.0	1,282.2
Student growth	0.340	-0.022	0.266	0.346	44.44	290.0	644.2
Autoregressive parameter (φ)	0.650	0.017	0.091	0.608	0.263	0.376	
ɛvel Two Variances							
Initial ability	1.000	0.084	0.612	0.853	7,999.9	33,671.2	$1.268*10^6$
Student growth	0.170	0.009	0.195	0.266	1,085.1	5,830.6	3.623*10
Autoregressive parameter (φ)	0.273	0.000	$4.322*10^{-5}$	0.001	0.003	0.001	0.0211

Item Response Theory Parameters

Overall, although estimated item response theory terms were biased when using the Kalman filter, the estimated and true values were highly correlated. None of the credible intervals overlapped with the true values, since the estimated values were a lot larger than the true values for every IRT parameter. Average estimated item discrimination was 1.507, compared to a true average of 0.9966, and average estimated item difficulty was 2.895, compared to a true average of 0.1884. On the other hand, actual item discrimination values were correlated with estimated values at 0.9849, while actual item difficulty values were correlated with their estimates at 0.9254.

Without using the Kalman filter, item response theory terms correlated very weakly to moderately with true values. Specifically, item discrimination estimates correlated with true values at 0.02730, while difficulty estimates correlated with true values at 0.4492. Seventy-five percent of the discrimination estimates were less than 0.000204, while the mean of the true discrimination terms was 0.9967. The median estimated difficulty term was 0.112, compared to the true median of 0.123; after removing an outlier difficulty of 3529.7, the estimated mean difficulty was 0.1211, and a true mean of 0.1884.

Level-one regression weights and ability terms.

Overall, Kalman filter estimation yielded biased estimates for individual-level initial ability estimates, individual-level growth terms, and individual-level autoregressive trends . However, individual ability and growth estimates were

moderately correlated with their true values. Under Kalman filter estimation, level-one autoregressive trends were underestimated and correlated with the true values at 0.04256. To elaborate, the mean estimated phi value was 0.08959, compared to an average true value of 0.5424. The underestimation is unsurprising given that the estimated level-two mean of that term is also underestimated. Similarly, the level-one growth terms were generally underestimated; the mean estimated level-one growth term was 0.2071, compared to an actual average value of 0.3437. On the other hand, the estimated level-one growth terms correlated with their true value at 0.6923. Finally, initial ability terms correlated with their true values at 0.8958, and were consistently overestimated; the mean estimated initial ability was 1.680, compared to the true mean of 0.04572. Also, under Kalman filter estimation, individual-level ability terms were consistently overestimated and correlated with their true values at only 0.2201. That is, the mean individual-level ability was estimated at 2.210, when the true value of this mean was 0.7391.

Estimating without using the Kalman filter resulted in estimates that were similarly correlated to the true values, but were generally overestimated. The exception is that the individual-level autocorrelation terms were slightly underestimated. The average estimated individual-level autocorrelation term was 0.3627, whereas the true value of this average was 0.5424. These terms correlated with each other at 0.1688. Next, the estimated individual-level growth terms were overestimated, with an average estimated value of 249, compared to the true value of 0.3437; however, estimated and true growth terms correlated at 0.4333. Finally, the individual-level initial ability terms were also overestimated, with an average

estimated value of 561.2, compared to a true average initial ability of 0.0457. Finally, this estimation method yielded individual-level ability estimates that correlated with the true values at 0.7046. The estimates were bimodal, with 980 estimates between 5621 and 5624, and the remaining estimates between -4.398 and 3.711.

Discussion

The fact that more estimates converged under the Kalman filter provides very modest evidence that Kalman filter estimation converges faster under the particular conditions simulated here. This may be purely due to the particular data that happened to be randomly generated, but in the absence of any other information, this thesis offers some small evidence that convergence is more efficient using the Kalman filter. However, taken as a whole,these results suggest that the Gibbs sampler, as described here, has an identification problem. Originally it was thought that fixing the first values of $\mathbf{a}$ and $\mathbf{b}$ would identify the model. However, both the item response theory terms and the initial ability estimates were overestimated, as a whole, suggesting that the model may not be identified at the second level. Future research may ameliorate this by fixing the mean of the initial ability estimate to be equal to zero.

What is potentially more troubling is that the variance of the autoregressive parameter is greatly underestimated, with the upper bound of a 95% credible interval very close to zero. It may be difficult to estimate the variance of this term since the autoregressive parameter is constrained to be between zero and one for stationarity purposes. Possibly the better solution would be to allow the sampler to obtain values outside this range, with the understanding that the growth model may or may not be stationary. Practically speaking, stationarity is mainly needed in the context of long-term forecasting, which may not be reasonable to do in Educational Psychology. For instance, it would be extremely difficult, if not theoretically implausible, to predict a student's college-level performance on

multivariate calculus from measurements taken on their arithmetic achievement fifteen years prior. Arguably, long-term forecasting of this nature is theoretically unreasonable for Educational Psychology.

Using the Kalman filter resulted in the convergence of more estimated terms. Particularly, estimating without using the Kalman filter resulted in nonconvergence of almost all of the autoregressive weights, and in more than one-quarter of the individual growth terms. For this particular simulated dataset, it appears that using the Kalman filter resulted in faster convergence of individual-level growth and autoregressive terms.

Not using the Kalman filter resulted in a variety of biased estimates, most notably the overestimation of variance and covariance terms. The extreme estimates for these variance terms may be due to the fact that many estimates have failed to converge without using the Kalman filter. As a result, these extreme estimates should be treated with caution, and those results do not necessarily provide evidence against not using the Kalman filter.

Limitations and future directions

The greatest limitation of this work is that only one dataset was simulated. It is possible that these findings may be a result of random variation in generating the dataset. The simplest extension of this work would be to simulate several datasets and determine the extent to which these findings are a result of simulation error.

Another major limitation is that certain parameters, namely a number of ability estimates, failed to converge under the Kalman filter. Results related to

those parameters that failed to converge should therefore be viewed with caution, or even ignored entirely by more cautious researchers. Particularly, the failure to obtain converged solutions for variance and covariance terms may impact the IRT parameters by impacting the scale of measurement. As such, the results for item difficulty and discrimination parameters must be interpreted with caution. That is because some part of these results may be due to the error of the Gibbs sampler. To address this, if greater computing power becomes available, it will be possible to continue running the model, since it is theoretically guaranteed to converge eventually. A larger simulation study may partially correct this problem; the lack of convergence here may be due to the generation of a particular dataset that is difficult to estimate.

A further limitation is that only a single set of parameters are investigated in this work. As such, these results do not generalize to situations where latent growth parameters are very different from the ones investigated here. For the future, a variety of variables, particularly the autoregression parameter, may be manipulated in a simulation study to evaluate the extent to which the Kalman filter reduces bias under a variety of conditions.

One of the advantages of second-order latent growth modeling is that item drift can be assessed over time, and the psychometric properties of different items at different time points may be evaluated. The present work is limited by simulating a single set of items which do not drift over time. Future research may extend this work by simulating item drift or including particularly easy or difficult time points at particular time points.

Also, the present study does not simulate any missing data, which is typically unrealistic for longitudinal datasets. Most large longitudinal datasets have participant attrition over time; also, participants may not be available for particular testing sessions. In theory, the Gibbs sampler as described here may be able to accommodate these missing data structures. That is, the sampler may estimate participant ability under missing data by using Equation 32, above, and use that term for the next iteration of the sampler. This relatively straightforward extension may be a fruitful topic for further research.

The simulations are also somewhat contrived in that it is assumed that every participant is measured at equal intervals. Occasionally participants are tested at different time points due to practical concerns. More typically in longitudinal datasets, the spacing between testing occasions is somewhat staggered from participant to participant. Future research may address this practical issue by simulating slightly different values of the time variable t in Equation 32, above, for each participant.

The Gibbs sampler described here may be further enhanced by including informative parameters for item response theory parameters. It may be the case that a previously calibrated measurement may be used on a different research sample to estimate a second-order latent growth model like the one described here. Using informative priors in this way is theoretically justifiable, since item calibration is thought to be person-free (Lord, 1952). These informative priors may be incorporated in a variety of different ways, which could be investigated in future research.

References

Barnes, G. M., Reifman, A. S., Farrell, M. P., & Dintcheff, B. A. (2000). The effects of parenting on the development of adolescent alcohol misuse: a Six-Wave latent growth model. *Journal of Marriage and Family, 62*(1), 175-186.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective* (Vol. 467). John Wiley & Sons.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008). *Time series analysis: Forecasting and control* (4th ed.). New York, NY: Wiley.

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics, 7*(4), 434-455.

Catts, H. W., Bridges, M. S., Little, T. D., & Tomblin, J. B. (2008). Reading achievement growth in children with language impairments. Journal of Speech, Language, and Hearing Research, 51(6), 1569-1579.

Chalmers, R. P. (2012). Mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software, 48*(6), 1-29.

Chan, D., & Schmitt, N. (2000). Interindividual differences in intraindividual changes in proactivity during organizational entry: a latent growth modeling approach to understanding newcomer adaptation. *Journal of Applied Psychology, 85*(2), 190.

Chatfield, C. (2004). *The analysis of time series* (6th ed.). Boca Raton, FL: Chapman-Hall/CRC Press.

Colder, C. R., et al., (2001). Identifying trajectories of adolescent smoking: an

application of latent growth mixture modeling. *Health Psychology*, *20*(2), 127.

Curran, P. J. (2000). A latent curve framework for studying developmental trajectories of adolescent substance abuse. In J. Rose, L. Chassin, C. Presson, & J. Sherman (Eds.), *Multivariate applications in substance use research* (pp. 1–42). Hillsdale, NJ: Lawrence Erlbaum Associates.

Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research, 38*(4), 529-569.

Curran, P. J., Stice, E., & Chassin, L. (1997). The relation between adolescent alcohol use and peer alcohol use: A Longitudinal Random Coefficients Model. *Journal of Consulting and Clinical Psychology, 65*(1), 130-140.

de Finnetti, B. (1974). *Theory of probability.* New York: Wiley.

Duncan, S. C., Duncan, T. E., Biglan, A., & Ary, D. (1998). Contributions of the social context to the development of adolescent substance use: a multivariate latent growth modeling approach. *Drug and alcohol dependence, 50*(1), 57-71.

Elliott, M. R., Gallo, J. J., Ten Have, T. R., Bogner, H. R., & Katz, I. R. (2005). Using a Bayesian latent growth curve model to identify trajectories of positive affect and negative events following myocardial infarction. *Biostatistics (Oxford, England)*, *6*(1), 119.

Ennet, S. T., & Bauman, K. E. (1994). The contribution of influence and selection to adolescent peer group homogeneity: The case of adolescent cigarette smoking. *Journal of Personality and Social Psychology, 67*(4), 653-663.

Ferrer, E., Balluerka, N., & Widaman, K. F. (2008). Factorial invariance and the specification of second-order growth models. *Methodology, 4*, 22–36.

Fleming, C. B., Mason, W. A., Mazza, J. J., Abbott, R. D., & Catalano, R. F. (2008). Latent growth modeling of the relationship between depressive symptoms and substance use during adolescence. *Psychology of Addictive Behaviors, 22*(2), 186.

Fox, J., & Glas, C. A. W. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika, 66*(2), 271-288.

Geiser, C., Keller, B. T., & Lockhart, G. (2013). First-versus second-order latent growth curve models: some insights from latent state-trait theory. *Structural equation modeling: a multidisciplinary journal*, *20*(3), 479-503.Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). *Bayesian data analysis* (Vol. 2). Boca Raton, FL, USA: Chapman & Hall/CRC.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevelhierarchical models* (Vol. 1). Cambridge: Cambridge University Press.

Gelman, A. and D. B. Rubin (1992) *Inference from iterative simulation using multiple sequences (with discussion)*. Statistical Science, 7:457-511.

Granger, C. W. J., & Morris, M. J. (1976). *Time Series Modeling and Interpretation.* Journal of the Royal Statistical Society, Series A (General), 139(2), 246-257.

Grimm, K. J., & Ram, N. (2009). Nonlinear growth models in M plus and SAS. *Structural Equation Modeling, 16*(4), 676-701.

Grimm, K. J., Ram, N., & Hamagami, F. (2011). Nonlinear growth curves in developmental research. *Child development, 82*(5), 1357-1371.

Hagenaars, J. A., & McCutcheon, A. L. (Eds.). (2002). *Applied latent class analysis*. Cambridge University Press.

Hamilton, J. D. (1994). *Time series analysis* (Vol. 2). Princeton: Princeton university press.

Hancock, G. R., Harring, J. R., & Lawrence, F. R. (2013). Using latent growth modeling to evaluate longitudinal change. In G. R. Hancock & R. O. Mueller (Eds.), *Structural Equation Modeling: A Second Course (2nd ed) (pp. 309-341).* Charlotte, NC: Information Age Publishing.

Hancock, G. R., Kuo, W. L., & Lawrence, F. R. (2001). An illustration of second-order latent growth models. *Structural Equation Modeling, 8*(3), 470-489.

Harring, J. R., Kohli, N., Silverman, R. D., & Speece, D. L. (2012). A second-order conditionally linear mixed effects model with observed and latent variable covariates. *Structural equation modeling: a multidisciplinary journal, 19*(1), 118-136.

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). A longitudinal study of

mathematical competencies in children with specific mathematics difficulties versus children with comorbid mathematics and reading difficulties. *Child development, 74*(3), 834-850.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of basic Engineering, 82*(1), 35-45.

Kline, R. B. (2011). Convergence of Structural Equation Modeling and Multilevel Modeling. In Williams, M., & Vogt, W. P. *Innovation in Social Research Methods* (pp. 562-589). Thousand Oaks, CA: Sage Publications Inc.

Kruschke, J. K. (2011). *Doing Bayesian Data Analysis: A Tutorial with R and BUGS.* Burlington, MA: Academic Press.

Kwok, O. M., West, S. G., & Green, S. B. (2007). The impact of misspecifying the within-subject covariance structure in multiwave longitudinal multilevel models: A Monte Carlo study. *Multivariate Behavioral Research, 42*(3), 557-592.

Leite, W. L. (2007). A Comparison of Latent Growth Models for Constructs Measured by Multiple Items. *Structural Equation Modeling, 14*(4): 581-610.

Lindley, D., V., & Novick, M. R. (1981). The role of exchangeability in inference. *Annals of statistics, 9*, 45-58.

Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013). Why the items versus parcels controversy needn't be one. *Psychological Methods, 18*(3), 285.

Lord, F. (1952). A theory of test scores. Psychometric monographs.

Murphy, D. L., Beretvas, S. N., & Pituch, K. A. (2011). The effects of autocorrelation on the curve-of-factors growth model. *Structural Equation Modeling: A Multidisciplinary Journal, 18*(3), 430-448.

Murphy, D. L., & Pituch, K. A. (2009). The performance of multilevel growth curve models under an autoregressive moving average process. *The Journal of Experimental Education, 77*(3), 255-284.

Muthén, B., & Asparouhov, T. (2011). Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. Handbook of advanced multilevel analysis, 15-40.

Petris, G., Petrone, S., & Campagnoli, P. (2009). Dynamic Linear Models with R. New York, NY: Springer Science+Business Media.

Rasbash, J., Steele, F., Browne, W. J., Goldstein, H., & Charlton, C. (2015). A User's Guide to MLwiN.

Roberts, G. O., & Smith, A. F. (1994). Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms. *Stochastic processes and their applications*, *49*(2), 207-216.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of statistics*, 34-58.

Sayer, A. G., & Cumsille, P. E. (2001). Second-order latent growth models.

Schumacker, R., & Lomax, R. G. (2010). Structural equation modeling. *NY. Routlege.*

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford university press.

Soares, T. M., Gonclaves, F. B., & Gamerman, D. (2009). An integrated Bayesian model for DIF analysis. *Journal of Educational and Behavioral Statistics, 34*, 348–377.

Steyer, R., Ferring, D., & Schmitt, M. J. (1992). States and traits in psychological assessment. *European Journal of Psychological Assessment*.

von Oerzen, T., Hertzog, C., Lindenberger, U., & Ghisletta, P. (2010). The effect of multiple indicators on the power to detect inter-individual differences in change. *British Journal of Mathematical and Statistical Psychology, 63*, 627-646.

Wright, B. D. (1967). Sample-free test calibration and person measurement. In Bloom, B. S. (Chair): *Invitational Conference on Testing Problems*. Princeton NJ: Educational Testing Service. Pp. 85-101.

Yuan, Y., & MacKinnon, D. P. (2009). Bayesian Mediation Analysis. *Psychological Methods, 14*, 301-322.